

Discrete Dispersion Models and their Tweedie Asymptotics

Bent Jørgensen

University of Southern Denmark,
Department of Mathematics and Computer Science,
Campusvej 55, DK-5230 Odense M, Denmark;
E-mail : bentj@stat.sdu.dk

Célestin C. Kokonendji

Université de Franche-Comté - UFR Sciences et Techniques
Laboratoire de Mathématiques de Besançon - UMR 6623 CNRS
16, route de Gray - 25030 Besançon cedex, France;
E-mail : celestin.kokonendji@univ-fcomte.fr

September 29, 2014

Abstract

We introduce a class of two-parameter discrete dispersion models, obtained by combining convolution with a factorial tilting operation, similar to exponential dispersion models which combine convolution and exponential tilting. The equidispersed Poisson model has a special place in this approach, whereas several overdispersed discrete distributions, such as the Neyman Type A, Pólya-Aeppli, negative binomial and Poisson-inverse Gaussian, turn out to be Poisson-Tweedie factorial dispersion models with power dispersion functions, analogous to ordinary Tweedie exponential dispersion models with power variance functions. Using the factorial cumulant generating function as tool, we introduce a dilation operation as a discrete analogue of scaling, generalizing binomial thinning. The Poisson-Tweedie factorial dispersion models are closed under dilation, which in turn leads to a Poisson-Tweedie asymptotic framework where Poisson-Tweedie models appear as dilation limits. This unifies many discrete convergence results and leads to Poisson and Hermite convergence results, similar to the law of large numbers and the central limit theorem, respectively. The dilation operator also leads to a duality transformation which in some cases transforms overdispersion into underdispersion and vice-versa. Many of the results have multivariate analogues, and in particular we consider a class of multivariate Poisson-Tweedie models, a multivariate notion of over- and underdispersion, and a multivariate zero-inflation index.

Keywords: factorial cumulant generating function; factorial tilting family; infinite dilatability; multivariate discrete distribution; over-/underdispersion; Poisson-Tweedie mixture

Mathematics Subject Classification: 60E10; 62E20; 62H05

1 Introduction

Given the plethora of discrete distributions available in the literature (Johnson *et al.*, 2005; Wimmer and Altmann, 1999), it is difficult to point, with conviction, to one or the other two-parameter discrete family as being especially suited for modelling count data phenomena such as over/underdispersion or zero-inflation/deflation. The central limit theorem leads to the normal distribution, which is continuous, and there are few general discrete asymptotic results available other than conventional Poisson convergence.

Echoing Tweedie (1984), who introduced the family of continuous Tweedie models now bearing his name, we should perhaps be looking for *an index which distinguishes between some important discrete distributions*. Ideally, such a class of discrete distributions should be justified by a general asymptotic result like the Tweedie convergence theorem of Jørgensen *et al.* (1994).

There are several problems that make the discrete case more difficult to handle than the continuous case. The first problem is that there are no immediate discrete analogues of location and scale transformations, which are crucial in the continuous case for handling scaling limits such as the central limit theorem. A second and related problem is that there are no obvious discrete analogues of standard continuous distributions such as the normal or gamma distributions. A third problem is that discrete natural exponential families (power-series distributions), while ubiquitous, tend to have much more complicated variance functions than in the continuous case.

An important step forward was taken by Steutel & van Harn (1979), who introduced the discrete analogue of positive stable distributions by using binomial thinning instead of scaling. The same technique has been used extensively for constructing discrete time-series models (e.g. Weiß, 2008). Recently, Harremoës *et al.* (2010) used binomial thinning to formulate an extended Poisson convergence theorem, which they called the "law of thin numbers", whereas Puig (2003) and Puig & Valero (2006, 2007) have characterized discrete distributions closed under convolution and binomial thinning.

In order to make further progress, we shall follow the footsteps of Jørgensen *et al.* (2010) and Jørgensen & Kokonendji (2011), who developed analogues of Tweedie asymptotics for extremes and geometric sums, respectively. These authors explored specialized versions of the cumulant generating function (CGF), and showed that each of the two corresponding analogues of the variance function are efficient characterization and convergence tools.

In the present paper we argue that the factorial cumulant generating function (FCGF) is the most suitable choice for handling the discrete case, along with the first two factorial cumulants, namely the mean and the dispersion. Firstly, the FCGF characterizes convolution additively. Secondly, we shall use the FCGF to generalize binomial thinning to a dilation operator, providing the discrete analogue of scaling. Thirdly, the *dispersion function*, which expresses the dispersion as a function of the mean, leads to a new discrete Poisson-Tweedie convergence theorem. Many known discrete distributions such as the Hermite, Neyman Type A, Pólya-Aeppli, binomial, negative binomial and Poisson-Inverse Gaussian distributions have power dispersion functions, and hence appear as limits in the corresponding regime of power asymptotics for dispersion functions. The corresponding power parameter is the index alluded to above.

The plan of the paper is to develop a new class of factorial dispersion models and Poisson-Tweedie mixtures as analogues of conventional exponential and Tweedie dispersion models, respectively, along the lines of Jørgensen (1997, Ch. 3-4). We review FCGFs and factorial cumulants in Section 2, we consider Poisson and Hermite convergence, and we consider the concept of infinite dilatability and its relation with Poisson mixtures. We introduce a new operation called the M-transformation, and show that in some cases it presents a duality between over- and underdispersion. In Section 3 we consider a new factorial tilting operation and introduce the class of factorial dispersion models and their dispersion functions. We show that the Poisson-Tweedie mixtures are factorial dispersion models and show that their dispersion functions are of power form. In Section 4 we present a general convergence theorem for dispersion functions (with proof given in Appendix B) and present the new Poisson-Tweedie convergence theorem and some examples. We consider the multivariate case in Section 5, where we discuss multivariate factorial cumulants and some of their properties, and consider multivariate over-, equi-, and underdispersion. We also introduce a new class of multivariate Poisson-Tweedie mixtures, which provides multivariate versions of many of the distributions mentioned above. Finally, Appendix A contains a summary of relevant results for exponential dispersion models.

2 Factorial cumulant generating functions

We begin by developing basic results for the FCGF and factorial cumulants, and use them to prove the law of thin numbers and Hermite convergence, which are discrete analogues of the law of large numbers and the central limit theorem, respectively. We define Poisson translation and dilation, and discuss infinite dilatability and its relation with Poisson mixtures. We also introduce the M-transformation and discuss its relation with over/underdispersion. Many results in the following deal with the *discrete case*, meaning non-negative integer-valued random variables, but unless otherwise indicated, results are valid for general random variables.

2.1 Cumulant generating functions

The ordinary cumulant generating function (CGF) for a random variable X is defined by

$$\kappa(s) = \kappa(s; X) = \log \mathbb{E}(e^{sX}) \text{ for } s \in \mathbb{R},$$

with effective domain $\text{dom}(\kappa) = \{s \in \mathbb{R} : \kappa(s) < \infty\}$. The CGF satisfies the linear transformation law

$$\kappa(t; aX + b) = \kappa(at; X) + bt, \quad (2.1)$$

which is crucial for asymptotic results like the law of large numbers and the central limit theorem.

To obtain a discrete analogue of (2.1), we consider the *factorial cumulant generating function* (FCGF) for X (Johnson *et al.*, 2005, p. 55), defined by

$$C(t) = C(t; X) = \log \mathbb{E} \left[(1+t)^X \right] = \kappa(\log(1+t); X) \text{ for } t > -1, \quad (2.2)$$

with effective domain $\text{dom}(C) = \{t > -1 : C(t) < \infty\} = \exp[\text{dom}(\kappa)] - 1$. We also note that C , like κ , characterizes convolution additively, i.e. for independent random variables X and Y we have

$$C(t; X + Y) = C(t; X) + C(t; Y). \quad (2.3)$$

The CGF κ is a real analytic convex function, and strictly convex unless X is degenerate. Hence, C is also real analytic, and the domain $\text{dom}(C)$, like $\text{dom}(\kappa)$, is an interval. The derivative $\dot{C}(t) = \dot{\kappa}(\log(1+t))/(1+t)$ has the same sign as $\dot{\kappa}(\log(1+t))$ on $\text{int}(\text{dom}(C))$. Hence, by the convexity of κ , the FCGF C is either monotone or u-shaped. Let \mathcal{K} denote the set of CGFs κ such that $\text{int}(\text{dom}(\kappa)) \neq \emptyset$, and let \mathcal{C} denote the corresponding set of FCGFs C of the form (2.2) with $\text{int}(\text{dom}(C)) \neq \emptyset$. In this case, either of the functions κ or C characterizes the distribution of X , and in equations like (2.2), we assume that equality holds in a neighbourhood of zero. From now on, CGF and FCGF refer to functions in \mathcal{K} and \mathcal{C} , respectively.

2.2 Dilation and Poisson translation

In order to obtain a discrete analogue of scaling, we define the *dilation* $c \cdot X$ of a random variable X by

$$C(t; c \cdot X) = C(ct; X), \quad (2.4)$$

for scalars $c > 0$ such that right-hand side of (2.4) is an FCGF. We say that X (or its distribution) is *infinitely dilatable* if the dilation $c \cdot X$ exists for any $c > 0$. For X discrete and $0 < c < 1$, the dilation corresponds to *binomial thinning*,

$$c \cdot X \stackrel{D}{=} \sum_{i=1}^X N_i, \quad (2.5)$$

where $\stackrel{D}{=}$ denotes equality in distribution, and where N_1, N_2, \dots is a sequence of i.i.d. Bernoulli random variables with probability parameter c , independent of X . To prove (2.5), we note that the Bernoulli FCGF is

$$C(t; N_1) = \log(1 + ct), \quad (2.6)$$

whereby

$$\log E[(1 + t)^{c \cdot X}] = \log E[(1 + ct)^X] = C(ct; X),$$

which implies (2.4). We note in passing, that the Bernoulli FCGF (2.6) is not infinitely dilatable, due to the constraint $c < 1$, whereas the geometric distribution with FCGF

$$C(t) = -\log(1 - \mu t) \quad (2.7)$$

is infinitely dilatable, since the domain for μ in (2.7) is \mathbb{R}_+ . The binomial thinning operator is an important tool for constructing discrete time series models, see e.g. Weiß (2008) and references therein.

A further extension of the dilation operator may be obtained by means of geometric compounding. Let us assume that, conditionally on a non-negative random variable X , we have a negative binomial FCGF $-X \log(1 - ct)$, where $c > 0$. The resulting negative binomial compound variable Y has FCGF

$$C(t; Y) = \log E[(1 - \mu t)^{-X}] = C(-ct; -X), \quad (2.8)$$

which is clearly infinitely dilatable. In the case where X is discrete, this is a geometric compounding of the form $N_1 + \dots + N_X$, where N_1, N_2, \dots is a sequence of i.i.d. geometric random variables with mean c , independent of X , corresponding to the negative binomial thinning of Ristić *et al.* (2009) and Barreto-Souza & Bourguignon (2014). For $c = 1$ and reversing the sign of t in (2.8) we obtain $C(t; -X) = C(-t; Y)$, providing a possible interpretation of the reflection operator.

The dilation operator satisfies the following associative property:

$$c_2 \cdot (c_1 \cdot X) \stackrel{D}{=} (c_1 c_2) \cdot X,$$

provided that the left-hand side of the equation exists. We also note the following distributive property of dilation for independent random variables X and Y ,

$$c \cdot (X + Y) \stackrel{D}{=} c \cdot X + c \cdot Y.$$

To obtain a discrete analogue of translation, we note that the Poisson distribution $\text{Po}(\mu)$ with mean $\mu \geq 0$ has FCGF

$$C(t; \text{Po}(\mu)) = \mu t \text{ for } t > -1, \quad (2.9)$$

including the degenerate case $\text{Po}(0) \equiv 0$. The Poisson FCGF is hence analogous to the CGF of a constant, and it is infinitely dilatable. We define the *Poisson translation* operator $\oplus \mu$ for $\mu \geq 0$ by convolution, i.e.

$$C(t; X \oplus \mu) = C(t; X) + \mu t.$$

As an example, we may write the stationary Poisson INAR(1) time series model (cf. McKenzie, 1985) in the following way:

$$X_t = c \cdot X_{t-1} \oplus [\lambda(1 - c)]$$

where $X_t \sim \text{Po}(\lambda)$ for $t = 0, 1, \dots$, and $0 < c < 1$. We also define the *Poisson subtraction* $\ominus \mu$ by

$$C(t; X \ominus \mu) = C(t; X) - \mu t \quad (2.10)$$

for values $\mu \geq 0$ such that the right-hand side of (2.10) is an FCGF. As an example, we consider the *Short* distribution (Johnson *et al.*, 2005, p. 419) with FCGF

$$C(t; X) = \mu_1 (e^{\phi t} - 1) + \mu_2 t$$

with $\mu_1, \mu_2, \phi > 0$. In this case, the Poisson subtraction (2.10) exists for $\mu \leq \mu_2$.

2.3 Factorial cumulants

When $0 \in \text{int}(\text{dom}(C))$, the derivatives $C^{(n)}(0) = C^{(n)}(0; X)$ are the *factorial cumulants* of X , which have many analogies with ordinary cumulants. The first factorial cumulant is the mean $E(X) = \dot{C}(0; X)$. The second factorial cumulant $S(X)$, defined by

$$S(X) = \ddot{C}(0; X) = \ddot{\kappa}(0; X) - \dot{\kappa}(0; X) = \text{Var}(X) - E(X),$$

is denoted the *dispersion* for X (compare with *Fisher's dispersion index* $D(X) = \text{Var}(X)/E(X)$). The dispersion is bounded below by the negative expectation,

$$-E(X) \leq S(X). \quad (2.11)$$

The dispersion $S(X)$ indicates *underdispersion* if $-E(X) \leq S(X) < 0$, *equidispersion* if $S(X) = 0$, and *overdispersion* if $S(X) > 0$, relative to the Poisson distribution.

The mean and dispersion satisfy the following transformation laws:

$$E(c \cdot X \oplus \mu) = cE(X) + \mu, \quad S(c \cdot X \oplus \mu) = c^2 S(X), \quad (2.12)$$

similar to the mean and variance of the linear transformation (2.1). Furthermore the n th factorial cumulant is homogeneous of degree n with respect to dilation, i.e. $C^{(n)}(0; c \cdot X) = c^n C^{(n)}(0; X)$. For general X and Y we obtain

$$S(X + Y) = S(X) + S(Y) + 2\text{Cov}(X, Y),$$

which follows from the corresponding formula for the variance. In particular, the dispersion is additive for uncorrelated random variables.

Applying the transformation laws (2.12) to the dilated variable $c \cdot X$, the inequality (2.11) implies that

$$-E(X) \leq cS(X).$$

Hence if $E(X)$ and $S(X)$ have opposite signs, the value of c is bounded from above or below, in which case X cannot be infinitely dilatable. In particular, if $E(X) > 0$ and X is underdispersed, then X is not infinitely dilatable, an example being the Bernoulli distribution (2.6).

Another index that may be obtained from the FCGF is the *zero-inflation index*, defined for a discrete random variable X by

$$\text{ZI}(X) = 1 + \frac{\log P(X = 0)}{E(X)} = 1 + \frac{C(-1; X)}{\dot{C}(0; X)}, \quad (2.13)$$

cf. Puig & Valero 2006, 2007). The index $\text{ZI}(X)$ indicates zero-inflation if $\text{ZI}(X) > 0$ and zero-deflation if $\text{ZI}(X) < 0$, relative to the Poisson distribution for which $\text{ZI}(X) = 0$. We consider a multivariate generalization of $\text{ZI}(X)$ in Section 5.1.

2.4 Poisson and Hermite convergence

We shall now present discrete analogues of the law of large numbers and the central limit theorem, obtained by exploring dilation and Poisson translation/subtraction. First we present a new proof of the *Law of Thin Numbers* due to Harremoës *et al.* (2010), which is a kind of Poisson law of large numbers for discrete distributions, in the sense that the Poisson distribution plays the role of a degenerate distribution. Our proof is based on the FCGF, whereas Harremoës *et al.* (2010) used direct methods in their proof.

We define the *dilation average* for i.i.d. sequence X_1, X_2, \dots by

$$\overline{X}_n = n^{-1} \cdot (X_1 + \dots + X_n). \quad (2.14)$$

For discrete random variables the dilation in (2.14) is defined by binomial thinning, because $n^{-1} \leq 1$.

Proposition 2.1 (Law of Thin Numbers). *Let X_1, X_2, \dots denote an i.i.d. sequence of discrete random variables with mean $\mu > 0$. Then the dilation average \overline{X}_n converges in distribution to a Poisson distribution:*

$$\overline{X}_n \xrightarrow{D} \text{Po}(\mu) \text{ as } n \rightarrow \infty. \quad (2.15)$$

Proof. By using the additive property (2.3) along with the definition of dilation, we obtain

$$\begin{aligned} C(t; \overline{X}_n) &= C(n^{-1}t; X_1 + \dots + X_n) \\ &= nC(n^{-1}t; X_1) = \mu t + O(n^{-1}), \end{aligned} \quad (2.16)$$

which converges to the Poisson FCGF as $n \rightarrow \infty$. Since $0 \in \text{int}(\text{dom}C(\cdot; \overline{X}_n))$, we conclude from Theorem 1 of Jensen & Nielsen (1997) that there exists a probability measure P such that the sequence of probability measures P_n corresponding to $C(\cdot; \overline{X}_n)$ converges weakly to P . It follows that the sequence $\exp C(e^s - 1; \overline{X}_n)$ converges to the moment generating function (MGF) of P for $e^s - 1 \in \text{dom}C(\cdot; \overline{X}_n)$, which in view of (5.4) implies that P is the Poisson distribution $\text{Po}(\mu)$, completing the proof. \square

For a constant integer variable n , the thinned variable $c \cdot n$ is binomial $\text{Bi}(c, n)$, corresponding to the special case $X_i \equiv 1$ of (2.14). We hence obtain the classical Poisson convergence theorem as a corollary, albeit in a rather terse notation.

Corollary 2.2 (Poisson Convergence). *For $\mu > 0$ we obtain*

$$\left(\frac{\mu}{n}\right) \cdot n \xrightarrow{D} \text{Po}(\mu) \text{ as } n \rightarrow \infty.$$

Before turning to Hermite convergence, we consider the Hermite distribution.

Example 2.1 (Hermite distribution). *The Hermite distribution (Kemp & Kemp, 1965), denoted $\text{PT}_0(\mu, \gamma)$ (conforming with the notation of Section 3.4) is defined by the FCGF*

$$C(t) = \frac{\gamma}{2}t^2 + \mu t, \quad (2.17)$$

where $\mu > 0$ is the mean, and γ is the dispersion, satisfying $0 < \gamma \leq \mu$. This restriction on the parameters follows from the corresponding log PGF

$$C(u - 1) = \frac{\gamma}{2}(u^2 - 1) + (\mu - \gamma)(u - 1),$$

whose coefficients γ and $\mu - \gamma$ must both be non-negative (Kemp & Kemp, 1965). The Hermite distribution is a discrete analogue of the normal distribution, in the sense that its third and higher order factorial cumulants are all zero. It is, however, rather different in nature from other discrete normal distributions such as those proposed by Kemp (1997) and Roy (2003). See Giles (2010) and Puig & Barquinero (2011) for applications of the Hermite distribution.

In order to obtain an analogue of the central limit theorem, we propose to use Poisson translation and dilation instead of centering and scaling. We consider the partial sum $S_n = X_1 + \dots + X_n$ based on i.i.d. discrete variables X_i with $E(X_i) = m > 0$ and $S(X_i) = \gamma > 0$. A formal analogy of the conventional centered and scaled partial sum of the form

$$Z_n = n^{-1/2} \cdot (S_n \ominus nm)$$

is, however, degenerate, because the centered sum $S_n \ominus nm$, is a non-negative variable with mean zero. By adding a constant $\mu \geq \gamma$, we obtain, formally,

$$\begin{aligned} Z_n \oplus \mu &= n^{-1/2} \cdot (S_n \ominus nm) \oplus \mu \\ &= n^{-1/2} \cdot \left[S_n \ominus (nm - n^{1/2}\mu) \right]. \end{aligned} \quad (2.18)$$

The constant being subtracted in (2.18) is now $nm - n^{-1/2}\mu < nm$, thereby avoiding the above degeneracy, at the cost of a constant Poisson translation. The expression (2.18) has the further advantage that the dilation by $n^{-1/2} \leq 1$ is again defined by binomial thinning. This leads us to the following analogue of the central limit theorem.

Proposition 2.3 (Hermite Convergence Theorem). *Let $S_n = X_1 + \dots + X_n$ denote the partial sum for an i.i.d. sequence of discrete random variables X_i with $E(X_1) = m > 0$ and $S(X_1) = \gamma > 0$. Define the translated standardized variable $Z_n(\mu)$ for $\mu \geq \gamma$ by*

$$Z_n(\mu) = n^{-1/2} \cdot \left[S_n \ominus \left(nm - n^{1/2}\mu \right) \right]$$

Then $Z_n(\mu)$ converges in distribution to the Hermite distribution $PT_0(\mu, \gamma)$ as $n \rightarrow \infty$.

Proof. The proof follows by expanding the FCGF of $Z_n(\mu)$ as follows:

$$C(t; Z_n(\mu)) = \mu t + \frac{\gamma}{2} t^2 + O\left(n^{-1/2}\right),$$

which shows that the Hermite FCGF (2.17) with $\mu \geq \gamma$ appears in the limit as $n \rightarrow \infty$. Using once more the results of Jensen & Nielsen (1997), we conclude that $Z_n(\mu)$ converges in distribution to the Hermite distribution $PT_0(\mu, \gamma)$. \square

A number of further convergence results will be considered in Section 4.

2.5 Poisson mixtures and infinite dilatability

We now discuss the relation between infinite dilatability and Poisson mixtures. If X is a non-negative random variable, and $a \geq 0$, we define the Poisson mixture $P(X; a)$ by the following conditional distribution:

$$P(X; a) | X \sim \text{Po}(aX),$$

see also Karlis & Xekalaki (2005). The corresponding conditional moment generating function (MGF) is

$$E\left(e^{sP(X; a)} | X\right) = \exp(aX(e^s - 1)).$$

Hence $P(X; a)$ has MGF

$$E\left(e^{sP(X; a)}\right) = E[\exp(aX(e^s - 1))] = e^{\kappa(a(e^s - 1); X)},$$

which implies that

$$C(t; P(X; a)) = \kappa(ta; X). \tag{2.19}$$

It follows that the factorial cumulants of $P(X; a)$ are obtained by scaling the ordinary cumulants for X . In particular, the first and second factorial cumulants are

$$E(P(X; a)) = aE(X) \text{ and } S(P(X; a)) = a^2 \text{Var}(X),$$

making $P(X; a)$ overdispersed, unless aX is degenerate.

In view of the scaling property $\kappa(ta; X) = \kappa(t; aX)$ of the CGF it follows from (2.19) that any Poisson mixture $P(X; a)$ is infinitely dilatable. The following result also contains the converse implication, similar to Theorem 3.1 of Jørgensen & Kokonendji (2011) for geometric infinite divisibility. This is an important prerequisite for our discussion of factorial tilting families in Section 3. We recall the definitions of the space \mathcal{K} of CGFs and the space \mathcal{C} of FCGFs, cf. Section 2.1.

Theorem 2.4. *Let the FCGF $C \in \mathcal{C}$ be given. Then the following conditions are equivalent:*

1. C is an infinitely dilatable FCGF;
2. $C(c \cdot) \in \mathcal{C}$ for any $c > 0$;
3. $C(c \cdot) \in \mathcal{K}$ for any $c > 0$;
4. C is the FCGF for a Poisson mixture $P(X; a)$.

Proof. 1. \Leftrightarrow 2.: This is the definition of infinite dilatability.

2. \Rightarrow 3.: Condition 2. implies that $C(n \cdot) \in \mathcal{C}$ for any integer n , so in view of (2.2) we find that $C(n(e^{sc/n} - 1))$ is a CGF for any $c > 0$. Letting $n \rightarrow \infty$ we obtain in the limit the function $C(c \cdot)$, which is hence a CGF for any $c > 0$, implying 3.

3. \Rightarrow 4.: By condition 3. we have that C is a CGF. The corresponding Poisson mixture (2.19) has FCGF $C(\cdot)$, which implies 4.

4. \Rightarrow 1.: This implication follows from (2.19) because $C(ct; P(X; a)) = C(t; P(X; ca))$ for any $c > 0$, which in turn implies that the Poisson mixture $P(X; a)$ is infinitely dilatable. This completes the proof. \square

Corollary 2.5. *Any infinitely dilatable FCGF $C \in \mathcal{C}$ is convex, and strictly convex except in the Poisson case.*

Proof. Since $C \in \mathcal{K}$ in the infinitely dilatable case, it follows that C is convex, and strictly convex unless $C(t) = \mu t$ for some $\mu \geq 0$, corresponding to the Poisson case (2.9). \square

For example, the geometric FCGF (2.7), being a Poisson mixture and hence infinitely dilatable, is strictly convex. Conversely, the Bernoulli FCGF (2.6), being strictly concave, is not infinitely dilatable, as we already know.

A Poisson mixture $P(X; a)$ may be expressed as a weighted Poisson distribution (Kokonendji *et al.* (2008); Kokonendji & Pérez-Casany (2012))

$$f_w(x; \mu) = \frac{w(x; \mu) \mu^x e^{-\mu}}{\mathbb{E}_\mu[w(X)] x!} \text{ for } x = 0, 1, \dots \quad (2.20)$$

where the weights have the form

$$w(x; \mu) = e^\mu (-1)^x \frac{d^x}{d\mu^x} \mathbb{E}(e^{-\mu X}).$$

The probabilities (2.20) may hence be calculated from the MGF for X .

2.6 The M-transformation and over/underdispersion

We now introduce a transformation that in some cases transforms underdispersion into overdispersion and vice-versa. Let us consider the reflected variable $-X$ with FCGF

$$C(t; -X) = \log \mathbb{E}[(1+t)^{-X}] = C\left(\frac{-t}{1+t}; X\right) \text{ for } t > -1.$$

Also recall that the dilation operator $X \mapsto c \cdot X$ is defined by $C(t; c \cdot X) = C(ct; X)$, for those $c > 0$ for which $C(ct; X)$ is an FCGF. The reflection and dilation operations do not commute, so that in general $c \cdot (-X)$ (if it exists) is different from $-c \cdot X$. We hence define the *M-transformation* by

$$X_a = c^{-1} \cdot [-c \cdot (-X)]$$

where $a = (1 - c)/c > -1$. The corresponding FCGF is

$$C(t; X_a) = C\left(\frac{t}{1 + at}; X\right).$$

The inverse M-transformation is defined by

$$X_a = -c^{-1} \cdot (-c \cdot X),$$

where $a = (c - 1)/c < 1$.

The first and second factorial cumulants for X_a are

$$E(X_a) = E(X) \quad \text{and} \quad S(X_a) = S(X) - 2aE(X).$$

It follows that the M-transformation may result in both overdispersion and underdispersion, depending on whether $2aE(X)$ is smaller or bigger than $S(X)$. The following example shows a case where the M-transformation maps overdispersion into underdispersion. Consider the negative binomial FCGF $-n \log(1 - \mu t)$ (with integer n). For $a = \mu \in (0, 1)$, the M-transformation maps the negative binomial FCGF into the binomial FCGF $n \log(1 + \mu t)$. The corresponding inverse transformation is obtained for $a = -\mu$.

The M-transformation hence provides one more tool in the study of over/underdispersion, see also Kokonendji *et al.* (2008), who used weighted Poisson distributions for this purpose. An application of the M-transformation is given in Section 3.6.

3 Factorial tilting and factorial dispersion models

We now introduce a factorial tilting operator, similar to exponential tilting, which leads to our main definitions of factorial tilting families and factorial dispersion models, providing discrete analogues of conventional exponential tilting and exponential dispersion models, respectively, as well as to the geometric dispersion models of Jørgensen & Kokonendji (2011). We also introduce the class of Poisson-Tweedie factorial dispersion models, which provide a parallel with the Tweedie class of exponential dispersion models, see Jørgensen (1997, Ch. 3–4). In the following we use the notation of Section 2.1.

3.1 Factorial tilting families

Consider the set $\overline{\mathcal{K}}$ of real analytic functions $K : \text{dom}(K) \rightarrow \mathbb{R}$ satisfying $0 \in \text{dom}(K)$ and $K(0) = 0$, where $\text{dom}(K)$ denotes the largest interval containing zero where K is analytic. We define the *tilting* of K by the amount $\theta \in \text{dom}(K)$ as the function $K_\theta : \text{dom}(K_\theta) \rightarrow \mathbb{R}$ given by

$$K_\theta(t) = K(\theta + t) - K(\theta) \text{ for } t \in \text{dom}(K_\theta) = \text{dom}(K) - \theta.$$

The tilting operator defines an equivalence relation on $\overline{\mathcal{K}}$. In particular, if $\kappa \in \mathcal{K}$, then κ_θ is the conventional exponential tilting of κ (cf. Jørgensen, 1997, p. 43). If we restrict the tilting operator to \mathcal{K} , the corresponding set of equivalence classes form the class of natural exponential families, i.e. CGF families of the form $\{\kappa_\theta \in \mathcal{K} : \theta \in \text{dom}(\kappa)\}$ for given $\kappa \in \mathcal{K}$. The corresponding natural exponential family has PDFs of the form

$$f(x; \theta) = g(x) \exp[\theta x - \kappa(\theta)] \text{ for } \theta \in \text{dom}(\kappa)$$

with respect to a suitable dominating measure, where g is the PDF corresponding to κ .

Let us now instead consider the restriction of the tilting operator to the class of FCGFs \mathcal{C} . We call this the *factorial tilting operator*. The corresponding set of equivalence classes in \mathcal{C} are called *factorial tilting families*, namely FCGF families of the form

$$\{C_\theta \in \mathcal{C} : \theta \in \text{dom}(C)\}$$

for given $C \in \mathcal{C}$. Note that when $C \in \mathcal{C}$, $\text{dom}(C)$ is restricted to the interval $t > -1$. The distribution with FCGF C_θ has mean $\mu = \dot{C}(\theta)$, and dispersion $\ddot{C}(\theta)$.

The factorial and exponential tilting operators turn out to be related by means of dilation. Thus, for given $\theta = e^\phi - 1 \in \text{dom}(C)$, and for $C(t) = \kappa(\log(1+t); X)$ as in (2.2), we obtain

$$\begin{aligned} C_\theta(t) &= C(\theta+t) - C(\theta) \\ &= \kappa(\log(1+\theta+t)) - \kappa(\log(1+\theta)) \\ &= \kappa\left(\log\left(1 + \frac{t}{1+\theta}\right) + \log(1+\theta)\right) - \kappa(\log(1+\theta)) \\ &= \kappa_\phi(\log(1+te^{-\phi})). \end{aligned} \tag{3.1}$$

The form (3.1) is an exponential tilting of κ , followed by a dilation. Conversely, the exponential tilting κ_ϕ corresponds to the FCGF

$$\kappa_\phi(\log(1+t)) = C_\theta(te^\phi),$$

which is a factorial tilting followed by a dilation.

In the special case of binomial thinning, we now derive the corresponding expression for the probability mass function (PMF) of a discrete model. If f is a given PMF, then the binomial thinning by c has PMF

$$f_c(x) = \sum_{i=x}^{\infty} f(i) \binom{i}{x} c^x (1-c)^{i-x}.$$

Now the exponential tilting by $\log(1+\theta)$ has density

$$f(x; \theta) = g(x) (1+\theta)^x e^{-C(\theta)}.$$

Now take $c = 1/(1+\theta)$ so that $1-c = \theta/(1+\theta)$. The binomial thinning by c is then

$$\begin{aligned} f_c(x; \theta) &= \sum_{i=x}^{\infty} g(i) (1+\theta)^i \binom{i}{x} (1+\theta)^{-x} [\theta/(1+\theta)]^{i-x} e^{-C(\theta)} \\ &= \sum_{i=x}^{\infty} g(i) \binom{i}{x} \theta^{i-x} e^{-C(\theta)}. \end{aligned}$$

This is the PMF of the binomial thinning of the exponential tilting for $\theta > 0$.

3.2 Dispersion functions

For a natural exponential family generated from the CGF κ , the variance function $V = \ddot{\kappa} \circ \dot{\kappa}^{-1}$ is known to be a useful characterization and convergence tool. We now introduce the dispersion function for factorial tilting families, and show that it has similar properties.

Let the FCGF $C = \kappa(\log(1+\cdot)) \in \mathcal{C}$ be given, and let $\{C_\theta : \theta \in \text{dom}(C)\}$ be the factorial tilting family generated by C . All factorial cumulants of C_θ are finite for $\theta \in \text{int}(\text{dom}(C))$, the first two being the mean

$$\mu = \dot{C}_\theta(0) = \dot{C}(\theta) = \dot{\kappa}(\log(\theta+1))/(\theta+1)$$

and the dispersion

$$\ddot{C}_\theta(0) = \ddot{C}(\theta) = \frac{\ddot{\kappa}(\log(\theta + 1)) - \dot{\kappa}(\log(\theta + 1))}{(\theta + 1)^2}.$$

Let $\Theta_0 \subseteq \text{dom}(C)$ be a non-degenerate interval where $\ddot{C}(\theta)$ has constant sign, such that $\dot{C}(\theta)$ is strictly monotone on Θ_0 , with $\mu = \dot{C}(\theta)$ belonging to the interval $\Psi_0 = \dot{C}(\Theta_0)$. Here we define μ by continuity at any end-point of Θ_0 contained in Θ_0 (Jørgensen, 1997, p. 46), allowing infinite values of μ , if necessary. We say that the family is *locally overdispersed* or *locally underdispersed* on Θ_0 , depending on the sign of $\ddot{C}(\theta)$. We may then parametrize the family locally by the mean μ , and we denote the corresponding family member by $\text{FT}(\mu)$. For a globally overdispersed or underdispersed family, we may parametrize the family globally by $\mu \in \Psi = \dot{C}(\text{dom}(C))$. We adopt the convention that for each $\mu \geq 0$, the Poisson distribution $\text{Po}(\mu)$ forms an equidispersed factorial tilting family.

Theorem 3.1. *Consider a locally overdispersed (underdispersed) factorial tilting family and define the local dispersion function $v : \Psi_0 \rightarrow \mathbb{R}$ by*

$$v(\mu) = \ddot{C} \circ \dot{C}^{-1}(\mu) \text{ for } \mu \in \Psi_0, \quad (3.2)$$

where v is defined by continuity at endpoints of Ψ_0 belonging to Ψ_0 , and where $v(\mu)$ is positive (negative) for all $\mu \in \Psi_0$. Then v characterizes the family among all factorial tilting families.

Proof. The proof is similar to the proof that a natural exponential family is characterized by its variance function (Jørgensen, 1997, p. 51). We first show that the dispersion function does not depend on the choice of the FCGF C representing the family. Thus, for given $\theta \in \Theta_0$, let us derive the local dispersion function corresponding to C_{θ_0} . For $t \in \text{dom}(C) - \theta$ we obtain $\dot{C}_\theta(t) = \dot{C}(\theta + t)$, so that $\dot{C}_\theta(\Theta_0 - \theta) = \dot{C}(\Theta_0) = \Psi_0$. The second derivative is $\ddot{C}_\theta(t) = \ddot{C}(\theta + t)$, and hence

$$\ddot{C}_\theta \circ \dot{C}_\theta^{-1}(\mu) = \ddot{C} \circ \dot{C}^{-1}(\mu) = v(\mu) \text{ for } \mu \in \Psi_0.$$

It follows that C_{θ_0} yields the same local dispersion function as C , so that v represents an intrinsic property of the family. To see that v characterizes the family among all factorial tilting families, we derive an inversion formula for v , again similar to the inversion formula for the variance function. If the FCGF C satisfies (3.2), then \dot{C}^{-1} satisfies the equation

$$\frac{d\dot{C}^{-1}}{d\mu}(\mu) = \frac{1}{\ddot{C} \circ \dot{C}^{-1}(\mu)} = \frac{1}{v(\mu)}.$$

For given v , the set of solutions to this equation are of the form $\dot{C}^{-1}(\mu) - \theta$, where $-\theta$ is an arbitrary constant. By solving the equation $t = \dot{C}^{-1}(\mu) - \theta$ with respect to μ we obtain $\mu = \dot{C}(\theta + t)$, and integration with respect to t in turn yields the function $C_\theta(t) = C(\theta + t) - C(\theta)$ satisfying the initial condition $C_\theta(0) = 0$. Since C_θ is an FCGF if and only if $\theta \in \text{dom}(C)$, we have thus recovered the factorial tilting family generated by C , as desired. \square

For a globally overdispersed or underdispersed family, we refer to v as simply the dispersion function. The fact that a factorial tilting family is characterized by the relations between its first two factorial cumulants provides an example of a family with finitely generated cumulants in the sense of Pistone & Wynn (1999). Khatri (1959) provides an early example of a characterization of this form. Note that positive v means that all members of the family are overdispersed, negative v means that all members are underdispersed, whereas zero v characterizes the Poisson family. The next result shows that many important factorial tilting families are Poisson mixtures, and hence overdispersed.

Proposition 3.2. *The family of Poisson mixtures (2.19) generated from a natural exponential family with variance function V yields an overdispersed factorial tilting family with dispersion function $v = V$.*

Proof. Consider the natural exponential family of CGFs κ_θ generated from the CGF $\kappa \in \mathcal{K}$. In view of (2.19), this family of CGFs is identical to the family of FCGFs for the corresponding Poisson mixtures, which hence form a factorial tilting family, and which is overdispersed due to the convexity of κ . The dispersion function of this family is identical to the variance function $\tilde{\kappa} \circ \kappa^{-1}$ of the natural exponential family. \square

Example 3.1 (Binomial and negative binomial distributions). *For each value of the convolution parameter $\lambda > 0$, the negative binomial FCGFs form a factorial tilting family,*

$$C_\theta(t) = -\lambda \log(1 - \mu t) \quad (3.3)$$

where $\mu = 1/(1 - \theta)$. The mean is $m = \lambda\mu$, and the dispersion function is $v(m) = \lambda^{-1}m^2$ for $m > 0$. Similarly, for each integer n , the binomial FCGFs also form a factorial tilting family,

$$C_\theta(t) = n \log(1 + \mu t) \quad (3.4)$$

where $\mu = 1/(1 + \theta)$. The mean is $m = \lambda\mu$, and the dispersion function is $v(m) = -n^{-1}m^2$ for $0 < m < n$.

Proposition 3.3. *The binomial, negative binomial and Poisson families are the only factorial tilting families that are closed under binomial thinning.*

Proof. Let $\text{FT}(\mu)$ denote the factorial tilting family with dispersion function $v(\mu)$. If the family is closed under binomial thinning, then $c \cdot \text{FT}(\mu) = \text{FT}(c\mu)$, and hence $c^2 v(\mu) = v(c\mu)$. Taking $m = c\mu$, this implies that $v(m)$ is either zero or proportional to m^2 . By Theorem 3.1 this, in turn, implies that $\text{FT}(\mu)$ is either one of the binomial, negative binomial or Poisson families. \square

In view of (3.1), we conclude that a factorial tilting family that is at the same time a natural exponential family must be closed under binomial thinning, and is hence either binomial or negative binomial. The Poisson natural exponential family is not included here, because each Poisson distribution $\text{Po}(\mu)$ is, on its own, a factorial tilting family.

3.3 Factorial dispersion models

We now introduce factorial dispersion models as two-parameter families of FCGFs obtained by combining the operations of factorial tilting and convolution/division. For given $C \in \mathcal{C}$ and $\lambda > 0$ we consider the following type of FCGF (*additive case*):

$$t \mapsto \lambda C_\theta(t) = \lambda C(\theta + t) - \lambda C(\theta) \quad (3.5)$$

for $\theta \in \Theta_0$. The second expression of (3.5) shows that the domain for (θ, λ) is a product set $\Theta_0 \times \Lambda$, with $\Lambda = \mathbb{R}_+$ if C is infinitely divisible. If C is not infinitely divisible, the domain Λ is a subset of \mathbb{R}_+ containing \mathbb{N} .

Like for exponential dispersion models, it is useful to consider as well the *reproductive case*, obtained by the dilation $\gamma = 1/\lambda$, which yields the FCGF

$$t \mapsto \gamma^{-1} C_\theta(\gamma t), \quad (3.6)$$

where the domain for the dispersion parameter γ is restricted to those values for which the dilation exists. We may parametrize a factorial tilting family locally (but not necessarily globally) by the mean $\mu = \dot{C}(\theta)$ of (3.6), in which case we denote the distributions corresponding to (3.5) and (3.6) by $\text{FD}^*(\mu, \lambda)$ and

$\text{FD}(\mu, \gamma)$, respectively. The dilation $\text{FD}(\mu, \gamma) = \gamma \cdot \text{FD}^*(\mu, \gamma^{-1})$ linking the two cases is called the *duality transformation*. The following table summarizes the two types of factorial dispersion models.

Type	Symbol	FCGF	Mean	Dispersion
Additive	$\text{FD}^*(\mu, \lambda)$	$\lambda C_\theta(t)$	$\lambda\mu$	$\lambda v(\mu)$
Reproductive	$\text{FD}(\mu, \gamma)$	$\gamma^{-1} C_\theta(\gamma t)$	μ	$\gamma v(\mu)$

We note in passing that the zero-inflation index $\text{ZI}(X) = 1 + C_\theta(-1; X)/\dot{C}_\theta(0; X)$ does not depend on the value of λ in the additive case, but only on μ .

For a factorial dispersion model generated by C , we refer to C and v as the unit FCGF and unit dispersion function, respectively. The additive form $\text{FD}(\mu, \gamma)$ is often useful because of its simple dispersion function $\gamma v(\mu)$, whereas $\text{FD}^*(\mu, \lambda)$, with mean $m = \lambda\mu$, say, has dispersion function $m \mapsto \lambda v(m/\lambda)$. An additive factorial tilting family $\text{FD}^*(\mu, \lambda)$ is closed under convolution,

$$\text{FD}^*(\mu, \lambda_1) + \text{FD}^*(\mu, \lambda_2) = \text{FD}^*(\mu, \lambda_1 + \lambda_2).$$

For Y_1, \dots, Y_n i.i.d. $\text{FD}(\mu, \gamma)$ then the dilation average $\bar{Y}_n = n^{-1} \cdot (Y_1 + \dots + Y_n)$ satisfies the following reproductive property:

$$\bar{Y}_n \sim \text{FD}(\mu, \gamma/n). \quad (3.7)$$

The additive binomial and negative binomial factorial dispersion models are apparent from (3.4) and (3.3), respectively. The corresponding reproductive FCGFs take the form

$$t \mapsto \pm \gamma^{-1} \log(1 \pm \gamma \mu t), \quad (3.8)$$

where $\gamma = 1/n$ or $\gamma = 1/\lambda$, respectively, which correspond to reparametrization of the two models in terms of the mean μ and the dispersion parameter γ .

3.4 Poisson-Tweedie mixtures and power dispersion functions

We have already introduced Poisson mixtures in Section 2.5, and we now consider the class of Poisson-Tweedie mixtures (Hougaard *et al.*, 1997; El-Shaarawi *et al.*, 2011), which are in many ways analogous to ordinary Tweedie models, and includes several well-known distributions as special cases.

Consider the Tweedie exponential dispersion model $\text{Tw}_p(\mu, \gamma)$, which has mean $\mu \in \Omega_p$, dispersion parameter $\gamma > 0$, and unit variance function

$$V(\mu) = \mu^p \quad \text{for } \mu \in \Omega_p, \quad (3.9)$$

where $p \notin (0, 1)$, $\Omega_0 = \mathbb{R}$, and $\Omega_p = \mathbb{R}_+$ for $p \neq 0$. The Poisson-Tweedie mixture $\text{PT}_p(\mu, \gamma)$ is defined as the Poisson mixture $\text{PT}_p(\mu, \gamma) = P(\text{Tw}_p(\mu, \gamma); 1)$. Here we require that $p \geq 1$, in order to make $\text{Tw}_p(\mu, \gamma)$ non-negative. For each $p \geq 1$, the Poisson-Tweedie mixture $Y \sim \text{PT}_p(\mu, \gamma)$ is an overdispersed factorial dispersion model with mean μ , unit dispersion function $v(\mu) = \mu^p$ defined by (3.9), and variance

$$\text{Var}(Y) = \mu + \gamma \mu^p. \quad (3.10)$$

The Poisson-Tweedie mixture $\text{PT}_p(\mu, \gamma)$ satisfies the following dilation property:

$$c \cdot \text{PT}_p(\mu, \gamma) = \text{PT}_p(c\mu, c^{2-p}\gamma) \quad \text{for } c > 0. \quad (3.11)$$

In the following, we use the notation $\text{PT}_p(\mu, \gamma)$ for any factorial dispersion model with power dispersion function, even if it is not a Poisson-Tweedie mixture.

The next theorem presents a characterization of factorial dispersion models that satisfy a dilation property like (3.11); similar to the characterization theorem for Tweedie exponential dispersion models

Table 1: The main types of factorial dispersion models with power dispersion functions.

Type	p	α
Hermite	$p = 0$	$\alpha = 2$
Poisson-binomial	$p = (n - 2)/(n - 1)$	$n = 3, 4, \dots$
Neyman Type A	$p = 1$	$\alpha = -\infty$
Poisson-negative binomial	$1 < p < 1$	$\alpha < 0$
Pólya-Aeppli	$p = 3/2$	$\alpha = -1$
Negative binomial/binomial	$p = 2$	$\alpha = 0$
Factorial discrete stable	$p > 2$	$0 < \alpha < 1$
Poisson-inverse Gaussian	$p = 3$	$\alpha = 1/2$

(Jørgensen, 1997, p. 128). Table 1 summarizes the main types of factorial dispersion models with power dispersion functions, including the Hermite and the Poisson-binomial distributions, which are not Poisson-Tweedie mixtures. Other values of p than those found in Table 1 are possible, as shown in Example 4.2 below.

Theorem 3.4. *Let $\text{FD}(\mu, \gamma)$ be a non-degenerate locally overdispersed factorial dispersion model satisfying $\inf \Psi_0 \leq 0$ or $\sup \Psi_0 = \infty$, such that for some $\gamma > 0$ and an interval of c -values*

$$c^{-1} \cdot \text{FD}(c\mu, \varphi_c \gamma) = \text{FD}(\mu, \gamma) \text{ for } \mu \in \Psi_0, \quad (3.12)$$

where φ_c is a positive function of c . Then $\text{FD}(\mu, \gamma)$ has power dispersion function proportional to μ^p for some $p \in \mathbb{R}$, and $\varphi_c = c^{2-p}$.

Proof. Calculating the dispersion on each side of (3.12) gives for an interval of c -values

$$c^{-2} \varphi_c \gamma v(c\mu) = \gamma V(\mu) \text{ for } \mu \in \Psi_0, \quad (3.13)$$

where v is the local unit dispersion function of $\text{FD}(\mu, \gamma)$. Taking, without loss of generality, $\mu = 1$ in (3.13) gives $\varphi_c = c^2 V(1)/V(c)$, which together with (3.13) implies that v satisfies the functional equation $v(1)v(c\mu) = v(c)v(\mu)$. This equation is equivalent to Cauchy's functional equation. By the continuity of v , the solutions to this equation are of the form $v(\mu) = \lambda \mu^p$ for some $p \in \mathbb{R}$ and $\lambda \neq 0$ because the family is non-degenerate (i.e. non-Poisson). This, in turn, implies that $\varphi_c = c^{2-p}$. In view of Theorem 3.1, $\text{FD}(\mu, \gamma)$ is hence a Poisson-Tweedie model in the case $p \geq 1$. For values of p less than 1, the model $\text{FD}(\mu, \gamma)$, if it exists, is not a Poisson-Tweedie mixture. \square

3.5 Discrete stable factorial dispersion models

The case $p > 2$ in Table 1 correspond to factorial dispersion models generated by discrete α -stable distributions with $\alpha \in (0, 1)$, where α is defined from the power parameter p by

$$\alpha = 1 + (1 - p)^{-1}, \quad (3.14)$$

with the convention that $\alpha = -\infty$ for $p = 1$ (Jørgensen, 1997, p. 131). In particular, the case $p = 3$ corresponds to the "Sichel" or Poisson-inverse Gaussian distributions with unit dispersion function $v(\mu) = \mu^3$; see Willmot (1987).

The discrete α -stable distribution with $\alpha \in (0, 1)$ was introduced by Steutel & van Harn, 1979), and corresponds to FCGFs proportional to

$$C^{(\alpha)}(t) = \frac{\alpha - 1}{\alpha} \left(\frac{t}{\alpha - 1} \right)^\alpha \text{ for } t/(\alpha - 1) > 0.$$

Factorial tilting and infinite division/convolution with power $\lambda > 0$ yield the Poisson-Tweedie mixture $\text{PT}_p^*(\mu, \lambda)$ (additive version) with FCGF

$$\lambda C_\theta^{(\alpha)}(t) = \lambda C^{(\alpha)}(\theta) [(1 + t/\theta)^\alpha - 1], \quad (3.15)$$

where the parameter μ is defined by

$$\mu = \dot{C}^{(\alpha)}(\theta) = \left(\frac{\theta}{\alpha - 1} \right)^{\alpha-1}. \quad (3.16)$$

An application of the duality transform then yields the Poisson-Tweedie mixture $\text{PT}_p(\mu, \gamma)$ with $\gamma = 1/\lambda$.

This construction of the Poisson-Tweedie mixtures is analogous to the construction of the Tweedie model $\text{Tw}_p(\mu, \gamma)$ as an exponential tilting of a positive α -stable distribution in the case $\alpha \in (0, 1)$. It is important to emphasize, however, that the above results could not have been easily obtained by means of exponential tilting. To illustrate this point, we note that the Poisson-inverse Gaussian mixture, when considered as an exponential dispersion model, has unit variance function given by

$$V(\mu) = \mu + \frac{\mu^3}{2} + \frac{\mu^2}{2} \sqrt{2 + \mu^2} \text{ for } \mu > 0,$$

as compared with the variance $\mu + \gamma\mu^3$ obtained from (3.10). For general $p \geq 1$, the Poisson-Tweedie exponential dispersion models have unit variance functions of the form

$$V(\mu) = \mu + \mu^p \exp[(2 - p)H(\mu)]$$

where $H(\mu)$ is implicitly defined (Kokonendji *et al.*, 2004; Jørgensen, 1997), in sharp contrast to (3.10). We also note that the so-called Hinde-Demétrio class of exponential dispersion models have unit variance functions of the form

$$V(\mu) = \mu + \mu^p,$$

but are not in general integer-valued (Kokonendji *et al.*, 2004).

3.6 Poisson-binomial and Poisson-negative binomial distributions

Consider the Poisson-negative binomial FCGF (Johnson *et al.*, 2005, p. 414), defined by

$$C(t) = \lambda \left[(1 - \mu t)^{-k} - 1 \right], \quad (3.17)$$

where $\lambda > 0$, which is essentially of the form (3.15) with $\alpha = -k < 0$, corresponding to $1 < p < 2$. The case $\lambda = -1$ ($p = 3/2$) is the Pólya-Aeppli distribution, and (3.17) is also known as a generalized Pólya-Aeppli distribution.

Similarly, let us consider the Poisson-binomial FCGF (Johnson *et al.*, 2005, p. 401), defined by

$$C(t) = \lambda [(1 + \mu t)^n - 1],$$

where $\lambda > 0$ and $n \in \mathbb{N}$. Up to a reparametrization, this FCGF is of the form (3.15) with $\alpha = n$, corresponding to $p = (n - 2)/(n - 1) \in (0, 1)$ for $n \geq 2$, which are not Poisson-Tweedie mixtures. The case $n = 1$ gives the Poisson distribution, whereas $n = 2$ gives the Hermite distribution. The Poisson-binomial distribution satisfies a binomial thinning property like (3.11) for $c \in (0, 1)$.

It is not immediately clear if there exist factorial dispersion models with $p \in (0, 1)$ corresponding to non-integer values of $\alpha > 2$. The following, formal considerations suggest that the answer to this question

may be affirmative. To this end, consider the M-transformation of the model $\text{PT}_p^*(\mu, \lambda)$ with $a = -1/\theta$, which has FCGF

$$\begin{aligned} C_\theta^{(\alpha)}\left(\frac{t}{1+at}\right) &= C^{(\alpha)}(\theta) \left[\left(1 + \frac{t/\theta}{1-t/\theta}\right)^\alpha - 1 \right] \\ &= C^{(\alpha)}(\theta) \left[(1-t/\theta)^{-\alpha} - 1 \right]. \end{aligned}$$

This FCGF is proportional to $C_{-\theta}^{(-\alpha)}(t)$, provided that the following ratio is positive:

$$\frac{C^{(\alpha)}(\theta)}{C^{(-\alpha)}(-\theta)} \propto (\alpha - 1) / (\alpha + 1),$$

which is the case for $|\alpha| > 1$. In particular, the set $1 < p < 4/3$ ($\alpha < -2$) is mapped onto the set $0 < p < 1$ ($\alpha > 2$). Similarly, the set $4/3 < p < 3/2$ ($-2 < \alpha < -1$) is mapped onto the set $p < 0$ ($1 < \alpha < 2$). The existence of the corresponding factorial dispersion models will be shown in Example 4.2 below.

3.7 Neyman Type A distribution

The Neyman Type A distribution $\text{PT}_1(\mu, \gamma)$ is a Poisson mixture of Poisson distributions, corresponding to the FCGF

$$C(t) = \gamma^{-1} \mu (e^{\gamma t} - 1),$$

see for example Dobbie & Welsh (2001) and Massé & Theodorescu (2005). The variance of $Y \sim \text{PT}_1(\mu, \gamma)$ is

$$\text{Var}(Y) = \mu(1 + \gamma),$$

which is special by not being asymptotic to μ near zero as is the case for Poisson-Tweedie mixtures with $p > 1$.

Like all reproductive Poisson-Tweedie mixtures, the parameter vector (μ, γ) is identifiable from the distribution $\text{PT}_1(\mu, \gamma)$, by means of the first two factorial cumulants μ and $\gamma v(\mu)$, similar to the case of reproductive exponential dispersion models. This is not, however, the case for the parameter (μ, λ) of the additive factorial dispersion model $\text{PT}_1^*(\mu, \lambda)$ with FCGF

$$C(t) = \lambda \mu (e^t - 1),$$

where only the mean $\lambda \mu$ is identifiable. The following result shows that this is essentially the only additive factorial dispersion model with this defect.

Theorem 3.5. *Consider a locally overdispersed or underdispersed additive factorial dispersion model $\text{FD}^*(\mu, \lambda)$. If the factorial tilting families $\text{FD}^*(\cdot, \lambda)$ are identical for an interval of λ -values, then $\text{FD}^*(\mu, \lambda)$ is a Neyman Type A family.*

Proof. We can assume, without loss of generality, that $(1, 1) \in \Psi_0 \times \Lambda$, the domain for (μ, λ) . Let v denote the unit dispersion function of $\text{FD}^*(\mu, \lambda)$. If the factorial tilting family $\text{FD}^*(\cdot, 1)$ is identical to the family $\text{FD}^*(\cdot, \lambda)$, then the two local dispersion functions are identical, i.e. $\lambda v(m/\lambda) = v(m)$, which for $m = 1$ implies $v(1/\lambda) = v(1)/\lambda$ for an interval of λ -values. We conclude that $v(\mu)$ is proportional to μ , which in view of Theorem 3.1 implies that $\text{FD}^*(\mu, \lambda)$ is a Neyman Type A family. \square

The situation is hence analogous to the case of additive exponential dispersion models, among which only the scaled Poisson family has this lack of identifiability (Jørgensen, 1997, p. 74).

4 Power asymptotics and Poisson-Tweedie convergence

We now consider power asymptotics for dispersion functions of factorial dispersion models, which proves convergence to distributions in the class of Poisson-Tweedie mixtures, similar to the Tweedie convergence theorem of Jørgensen *et al.* (1994), see also Jørgensen (1997, Ch. 4). This approach provides a unified method of proof for a range of different convergence results for discrete distributions, many of which are new.

4.1 Convergence of dispersion functions

We first present a general convergence theorem for factorial tilting families, which is used for proving the Poisson-Tweedie convergence theorem below (Theorem 4.2). The result is similar to the Mora (1990) convergence theorem for variance functions (Jørgensen, 1997, p. 54), which says that convergence of a sequence of variance functions, when the convergence is uniform on compact sets, implies weak convergence of the corresponding sequence of natural exponential families.

Theorem 4.1. *Let $\{\text{FT}_n(\mu) : n = 1, 2, \dots\}$ denote a sequence of locally overdispersed or underdispersed factorial tilting families having local dispersion functions v_n with domains Ψ_n . Suppose that*

1. $\bigcap_{n=1}^{\infty} \Psi_n$ contains a non-empty interval Ψ_0 ;
2. $\lim_{n \rightarrow \infty} v_n(\mu) = v(\mu)$ exists uniformly on compact subsets of $\text{int}\Psi_0$;
3. $v(\mu) \neq 0$ for all $\mu \in \text{int}\Psi_0$ or $v(\mu) = 0$ for all $\mu \in \text{int}\Psi_0$.

In the case $v(\mu) \neq 0$, there exists a factorial tilting family $\text{FT}(\mu)$ whose local dispersion function coincides with v on $\text{int}\Psi_0$, such that for each μ in $\text{int}\Psi_0$ the sequence of distributions $\text{FT}_n(\mu)$ converges weakly to $\text{FT}(\mu)$. In the case $v(\mu) = 0$, $\text{FT}_n(\mu)$ converges weakly for each μ in $\text{int}\Psi_0$ to the Poisson distribution $\text{Po}(\mu)$.

The proof of Theorem 4.1, which is given in Appendix B, is similar to the proof by Mora (1990), see also Jørgensen (1997, p. 54). The case of convergence to a zero dispersion function follows the same line of proof as in Jørgensen & Kokonendji (2011) for geometric dispersion models.

We now use Theorem 4.1 to give a new proof of the Poisson law of thin numbers (Proposition 2.1). Let us first note that a locally overdispersed or underdispersed reproductive factorial dispersion model $\text{FD}(\mu, \gamma)$ has local dispersion function of the form $\gamma v(\mu)$, which goes to zero as $\gamma \downarrow 0$. It is easy to show that the limit exists uniformly on compact subsets of Ψ_0 . By Theorem 4.1 this implies

$$\text{FD}(\mu, \gamma) \xrightarrow{D} \text{Po}(\mu) \text{ as } \gamma \downarrow 0. \quad (4.1)$$

This applies, in particular, to all Poisson-Tweedie mixtures and power dispersion function models $\text{PT}_p(\mu, \gamma)$. The result implies that all factorial dispersion models resemble the Poisson distribution for small dispersion, irrespective of their origin. Furthermore, consider the dilation average \overline{Y}_n based on Y_1, \dots, Y_n i.i.d. from $\text{FD}(\mu, \gamma)$, which, by (3.7), has distribution $\overline{Y}_n \sim \text{FD}(\mu, \gamma/n)$. By (4.1) this implies that

$$\overline{Y}_n \xrightarrow{D} \text{Po}(\mu) \text{ as } n \rightarrow \infty.$$

We have hence obtained a new proof of the law of thin numbers. This Poisson convergence result is analogous to the exponential convergence result for geometric dispersion models of Jørgensen & Kokonendji (2011).

4.2 Power asymptotics

To motivate the next Poisson-Tweedie convergence theorem, let us rewrite the dilation result (3.11) in the form of a fixed point

$$c^{-1} \cdot \text{PT}_p(c\mu, c^{2-p}\gamma) = \text{PT}_p(\mu, \gamma).$$

The next theorem shows that this fixed point has a domain of attraction characterized by a power asymptotic dispersion function. The theorem is analogous to the Tweedie convergence theorem for exponential dispersion models (Jørgensen, 1997, pp. 148–149) and to similar convergence results for extreme and geometric dispersion models (Jørgensen *et al.*, 2010; Jørgensen & Kokonendji, 2011).

Theorem 4.2. *Let $\text{FD}(\mu, \gamma)$ denote a locally overdispersed or underdispersed factorial dispersion model with unit dispersion function v on Ψ_0 , such that either $\inf \Psi_0 \leq 0$ or $\sup \Psi_0 = \infty$. Assume that for some $p \in \mathbb{R}$ the unit dispersion function satisfies $v(\mu) \sim c_0 \mu^p$ as either $\mu \downarrow 0$ or $\mu \rightarrow \infty$. Then for each $\mu \in \Omega_p$*

$$c^{-1} \cdot \text{FD}(c\mu, c^{2-p}\gamma) \xrightarrow{D} \text{PT}_p(\mu, \gamma c_0) \quad \text{as } c \downarrow 0 \text{ or } c \rightarrow \infty, \quad (4.2)$$

respectively. In the case $c \downarrow 0$, the model $\text{FD}(\mu, \gamma)$ is required to be infinitely dilatable, and if $c^{2-p} \rightarrow \infty$ the model is required to be infinitely divisible.

Proof. Without loss of generality we may take $c_0 = 1$. We first note that for each given value of γ and c , the left-hand side of (4.2) is a factorial tilting family with mean μ , provided that c is small (large) enough for $c\mu$ to belong to Ψ_0 . The corresponding dispersion function satisfies

$$c^{-2} c^{2-p} \gamma v(c\mu) \rightarrow \gamma \mu^p \text{ as } c \downarrow 0 \text{ or } c \rightarrow \infty,$$

respectively, and hence converges to the dispersion function of $\text{PT}_p(\mu, \gamma)$. To show that the convergence is uniform in μ on compact subsets of Ω_p , let us consider the case where $c \downarrow 0$ (the proof is similar in the case $c \rightarrow \infty$). Let $0 < M_1 \leq \mu \leq M_2 < \infty$ and $\epsilon > 0$ be given, and let c be small enough to make

$$\left| \frac{v(c\mu)}{(c\mu)^p} - 1 \right| < \epsilon$$

for all $\mu \leq M_2$. Then

$$\left| \frac{v(c\mu)}{c^p} - \mu^p \right| = \mu^p \left| \frac{v(c\mu)}{(c\mu)^p} - 1 \right| \leq (M_1^p + M_2^p) \epsilon,$$

which shows that the convergence is uniform on the compact interval $M_1 \leq \mu \leq M_2$. The result (4.2) now follows from Theorem 4.1. \square

Many factorial dispersion models have power asymptotic dispersion functions, and are hence asymptotically similar to Poisson-Tweedie mixtures. Thus, under the hypothesis of Theorem 4.2, the dilation property (3.11) for $\text{PT}_p(\mu, \gamma)$ implies the following distribution approximation:

$$\text{FD}(c\mu, c^{2-p}\gamma) \dot{\sim} \text{PT}_p(c\mu, c^{2-p}\gamma c_0) \quad (4.3)$$

for c small or c large, respectively. In view of the fact that any FCGF belongs to some factorial dispersion model (namely the model generated by the FCGF itself), many factorial dispersion models may be approximated by Poisson-Tweedie models in this way.

Example 4.1 (Discrete Linnik distribution). *The discrete Linnik distribution is defined by the FCGF*

$$C(t) = -b \log [1 + c(-t)^\alpha] \quad (4.4)$$

where $0 < \alpha < 1$ and $b, c > 0$ (Johnson et al. , 2005, p. 497). This distribution is hence infinitely dilatable as well as infinitely divisible. The asymptotic behaviour of $C(t)$ as $t \uparrow 0$ is

$$C(t) \sim -\lambda c(-t)^\alpha,$$

which, in turn, implies that the unit dispersion function is power asymptotic at infinity,

$$v(\mu) \sim c_0 \mu^p \text{ as } \mu \rightarrow \infty$$

for some $c_0 > 0$, where $p > 2$ is related to $\alpha \in (0, 1)$ by (3.14). It follows that the factorial dispersion model generated by $C(t)$ satisfies (4.2) as $c \rightarrow \infty$.

To connect the result (4.2) with large sample theory, let $\bar{Y}_n \sim \text{FD}(\mu, \gamma/n)$ denote the dilation average of an i.i.d. sample from the distribution $\text{FD}(\mu, \gamma)$ (cf. Eq. (3.7)). Then for $p \neq 2$ we may rewrite (4.2) as follows (taking $c^{2-p} = 1/n$):

$$n^{-1/(p-2)} \cdot \text{FD}(n^{1/(p-2)}\mu, \gamma/n) \xrightarrow{D} \text{PT}_p(\mu, \gamma c_0) \quad \text{as } n \rightarrow \infty, \quad (4.5)$$

so the scaled and factorially tilted dilation average \bar{Y}_n converges to a Poisson-Tweedie model. We interpret this result via (4.3) as saying that a system subject to independent $\text{FD}(\mu, \gamma)$ -distributed shocks will eventually settle in what may be called a Poisson-Tweedie equilibrium.

Alternatively, let us consider the case where $\bar{Y}_n \sim \text{FD}(\mu, \gamma)$ is the dilation average of an i.i.d. sample from the distribution $\text{FD}(\mu, n\gamma)$, which requires that the model $\text{FD}(\mu, \gamma)$ be infinitely divisible. Then for $p \neq 2$ we may rewrite (4.2) as follows (taking $c^{2-p} = n$):

$$n^{-1/(2-p)} \cdot \text{FD}(n^{1/(2-p)}\mu, \gamma n) \xrightarrow{D} \text{PT}_p(\mu, \gamma c_0) \quad \text{as } n \rightarrow \infty. \quad (4.6)$$

We interpret the result (4.6) as saying that the scaled and factorially tilted component $\text{FD}(\mu, n\gamma)$ converges to a Poisson-Tweedie model. The main feature of (4.6) is that the signs of the powers of n are reversed compared with (4.5).

The results (4.5) and (4.6) both require $p \neq 2$, which together with Proposition 3.3 highlights the special role of the binomial and negative binomial distributions in Poisson-Tweedie asymptotics, as we shall now see.

4.3 Binomial and negative binomial convergence

We now discuss the power asymptotics of Theorem 4.2 in the case $p = 2$ ($\alpha = 0$). We first note that the dilation property (3.12) for the negative binomial distribution $\text{PT}_2(\mu, \gamma)$ takes the form

$$c^{-1} \cdot \text{PT}_2(c\mu, \gamma) = \text{PT}_2(\mu, \gamma) \text{ for } c > 0,$$

for all $\mu > 0$ and $\gamma > 0$. Suppose that the locally overdispersed factorial dispersion model $\text{FD}(\mu, \gamma)$ with mean domain Ψ_0 is such that either $\inf \Psi_0 \leq 0$ or $\sup \Psi_0 = \infty$, and assume that the unit dispersion function satisfies $v(\mu) \sim \mu^2$ as $\mu \downarrow 0$ or $\mu \rightarrow \infty$, respectively. The corresponding version of (4.2) is then

$$c^{-1} \cdot \text{FD}(c\mu, \gamma) \xrightarrow{D} \text{PT}_2(\mu, \gamma) \text{ as } c \downarrow 0 \text{ or } c \rightarrow \infty, \quad (4.7)$$

respectively, for all $\mu, \gamma > 0$. The result (4.7) implies the following negative binomial approximation:

$$\text{FD}(c\mu, \gamma) \dot{\sim} \text{PT}_2(c\mu, \gamma) \quad \text{as } c \downarrow 0 \text{ or } \rightarrow \infty,$$

respectively. The result does not involve a large sample in any sense, but instead applies as the mean $c\mu$ goes to the boundary of the parameter space.

An example is the factorial dispersion model $\text{FD}(\mu, \gamma)$ generated by the discrete Linnik distribution (4.4). The corresponding unit dispersion function behaves as $c_0\mu^2$ as $\mu \downarrow 0$ for some $c_0 > 0$. It follows that (4.7) is satisfied as $c \downarrow 0$.

We now turn to the binomial distribution $\text{Bi}(\mu, n)$, which is a factorial dispersion model on additive form. The binomial distribution satisfies the following thinning property:

$$c^{-1} \cdot \text{Bi}(c\mu, n) = \text{Bi}(\mu, n) \text{ for } 0 < c < 1.$$

As a result, suppose that the underdispersed additive factorial dispersion model $\text{FD}^*(\mu, n)$ has unit dispersion function satisfying $v(\mu) \sim -\mu^2$ as $\mu \downarrow 0$. Then

$$c^{-1} \cdot \text{FD}^*(c\mu, n) \xrightarrow{D} \text{Bi}(\mu, n) \quad \text{as } c \downarrow 0. \quad (4.8)$$

Example 4.2 (COM-Poisson distribution). *Consider the COM-Poisson distribution (cf. Shmueli et al., 2005) with PMF*

$$P(X = x) = \frac{\lambda^x}{(x!)^\nu Z(\lambda, \nu)} \text{ for } x = 0, 1, \dots, \quad (4.9)$$

where $\lambda > 0$, $\nu \geq 0$, and $Z(\lambda, \nu)$ is a normalizing constant. This family provides useful illustrations of several of the above convergence results. The FCGF of (4.9) is

$$C(t) = \log \frac{Z(\lambda(1+t), \nu)}{Z(\lambda, \nu)}, \quad (4.10)$$

and the first two factorial cumulants of the local factorial tilting family $\text{FT}(\mu)$ generated by (4.9) (for given value of (λ, ν)) are

$$\mu = \lambda \frac{\dot{Z}(\lambda(1+\theta), \nu)}{Z(\lambda(1+\theta), \nu)} \text{ and } \lambda^2 \frac{\ddot{Z}(\lambda(1+\theta), \nu)}{Z(\lambda(1+\theta), \nu)} - \mu^2,$$

respectively, where dots denote derivatives of $Z(\cdot, \nu)$. These results confirm, for $\theta = 0$, known results for the mean and variance of the COM-Poisson distribution. It is well known that the COM-Poisson converges to the Bernoulli distribution $\text{Bi}(\lambda/(1+\lambda), 1)$ as $\nu \rightarrow \infty$. A similar result using the above binomial convergence result is obtained by using the following asymptotic relation: $\log Z(\lambda(1+\theta), \nu) \sim \log[1 + \lambda(1+\theta)]$ as $\theta \downarrow -1$, which implies that the local dispersion function of the factorial tilting family $\text{FT}(\mu)$ defined by (4.10) satisfies $v(\mu) \sim -\mu^2$ as $\mu \downarrow 0$. Using (4.8), we hence obtain the following Bernoulli convergence:

$$c^{-1} \cdot \text{FT}(c\mu) \xrightarrow{D} \text{Bi}(\mu, 1) \quad \text{as } c \downarrow 0.$$

Turning now to the question of infinite divisibility, we note that Kokonendji et al. (2008) argued that the COM-Poisson is a weighted Poisson distribution of the form (2.20) with a logconvex (logconcave) weight function for $0 < \nu < 1$ ($\nu > 1$), and is hence overdispersed (underdispersed) with respect to the Poisson case $\nu = 1$. Following Kokonendji et al. (2008), we may further argue that the distribution is infinitely divisible in the logconvex case $0 < \nu < 1$, whereas for $\nu > 1$, we have a discrete underdispersed distribution, which cannot be infinitely divisible. In order to apply Theorem 4.2, we need the following asymptotic expansion, gleaned from Sellers et al. (2012),

$$C(t) \sim \alpha^{-1} \lambda^\alpha (1+t)^\alpha - \frac{1-\alpha}{2} \log \lambda (1+t) \text{ as } t \rightarrow \infty, \quad (4.11)$$

where $\alpha = 1/\nu > 0$; the first term of the expansion being the leading term. Let p be related to α via (3.14). In the overdispersed case $\alpha > 1$ ($p < 1$), the result (4.11) implies that the dispersion function is power asymptotic,

$$v(\mu) \sim (\alpha - 1) \lambda^{\alpha/(\alpha-1)} \mu^p \quad (4.12)$$

as $\mu \rightarrow \infty$. The overdispersed factorial dispersion model $\text{FD}(\mu, \gamma)$ generated by (4.9) hence satisfies (4.2) as $c \rightarrow \infty$. This result is remarkable, in that it proves the existence of the power dispersion model $\text{PT}_p(\mu, \gamma)$ in the cases $p < 0$ and $0 < p < 1$ (cf. Theorem 3.4 and Table 1 above), because Theorem 4.1 implies the existence of the factorial dispersion model corresponding to the limiting local dispersion function. In the underdispersed case $0 < \alpha < 1$ ($p > 2$) we find that (4.12) is now satisfied as $\mu \downarrow 0$, but with a negative coefficient for μ^p . In this case, however, Theorem 4.2 does not apply, because the case $c \downarrow 0$ in (4.2) requires infinite divisibility, which we do not have in the underdispersed case, nor do we seem to have infinite dilatability. These results, while interesting on their own, are to some extent tangential to the COM-Poisson distribution itself, because the factorial dispersion model $\text{FD}(\mu, \gamma)$ is not contained in the COM-Poisson family.

4.4 Neyman Type A convergence

A new result that emerges from Poisson-Tweedie asymptotics is convergence to the Neyman Type A distribution $\text{PT}_1(\mu, \gamma)$ (cf. Section 3.7). This is the case, in particular, for a certain type of Poisson mixtures.

Let us assume that the exponential dispersion model $\text{ED}(\mu, \gamma)$ has unit variance function satisfying $V(\mu) \sim \mu$ as $\mu \downarrow 0$ or $\mu \rightarrow \infty$. From Proposition 3.2 we obtain that the Poisson mixture $P(\text{ED}(\mu, \gamma); 1)$ is a factorial dispersion model with unit dispersion function $v = V$. It then follows from Theorem 4.2, that the corresponding tilted and dilated model converges to the Neyman Type A distribution,

$$c^{-1} \cdot \text{FD}(c\mu, c\gamma) \xrightarrow{D} \text{PT}_1(\mu, \gamma) \quad \text{as } c \downarrow 0 \text{ or } \rightarrow \infty,$$

respectively.

Let $\text{ED}^*(\mu, \lambda)$ denote an additive exponential dispersion model generated by a distribution with an atom at zero, and such that $(0, 1)$ is the largest interval starting at zero with zero probability. This may happen if the distributions has support \mathbb{N}_0 , but the distribution need not necessarily be discrete as long as there is an atom at zero and positive probability at 1 or starting at 1. Then we know from Jørgensen *et al.* (1994) that the unit variance function satisfies $V(\mu) \sim \mu$ as $\mu \downarrow 0$. The corresponding exponential dispersion model is $\gamma \text{ED}^*(\mu, \gamma^{-1})$. Hence, let us consider the factorial dispersion model $\text{FD}(\mu, \gamma)$ defined by the Poisson mixture $P(\gamma \text{ED}^*(\mu, \gamma^{-1}); 1)$. Then we have the following large-sample convergence result:

$$n \cdot \text{FD}(\mu/n, \gamma/n) \xrightarrow{D} \text{PT}_1(\mu, \gamma) \quad \text{as } n \rightarrow \infty. \quad (4.13)$$

Here, using (3.7), the left-hand side of (4.13) may be interpreted as the sum of n i.i.d. random variables with distribution $\text{FD}(\mu/n, \gamma)$.

4.5 Hermite convergence revisited

The Hermite distribution $\text{PT}_0(\mu, \gamma)$ of Example 2.1 has power dispersion function with $p = 0$ ($\alpha = 2$), although it is not a Poisson-Tweedie mixture. In Section 2.4, we have already considered a type of Hermite convergence similar to the central limit theorem. We now consider Hermite convergence based on Theorem 4.2.

It is important to keep in mind that the parameters of the Hermite distribution must satisfy $0 < \gamma \leq \mu$. The dilation property (3.11) hence takes the following form:

$$c \cdot \text{PT}_0(\mu, \gamma) = \text{PT}_0(c\mu, c^2\gamma) \text{ for } c < 1,$$

where the restriction $c < 1$ ensures that the transformed parameters satisfy the condition. As a consequence, only the case $c \downarrow 0$ of (4.2) is available in the Hermite case.

We first note that the local dispersion function of any locally overdispersed factorial dispersion model $\text{FD}(\mu, \gamma)$ with $0 \in \text{int}\Psi_0$ satisfies $v(\mu) \sim v(0) > 0$ as $\mu \downarrow 0$. By Theorem 4.2, and using the form (4.5) we obtain convergence to the Hermite distribution,

$$n^{1/2} \cdot \text{FD}(n^{-1/2}\mu, \gamma/n) \xrightarrow{D} \text{PT}_0(\mu, \gamma v(0)) \text{ as } n \rightarrow \infty, \quad (4.14)$$

for each $\mu > 0$, provided that $\gamma v(0) \leq \mu$. By referring once more to Eq. (3.7), we note that the left-hand side of (4.14) involves a dilated and factorially tilted dilation average of n i.i.d. variables from the distribution $\text{FD}(n^{-1/2}\mu, \gamma)$.

The condition that $\gamma v(0) \leq \mu$ may be alleviated by means of Poisson translation, similar to the procedure of Section 2.4. This leads to the following result:

$$n^{1/2} \cdot \left[\text{FD}(\mu_0 + n^{-1/2}\mu, \gamma/n) \ominus \mu_0 \right] \xrightarrow{D} \text{PT}_0(\mu, \gamma v(\mu_0)) \text{ as } n \rightarrow \infty,$$

for $\mu_0 \in \Psi_0$, provided that μ is chosen such that $\gamma v(\mu_0) \leq \mu$. Here \ominus denotes Poisson subtraction, as defined by (2.10).

5 Multivariate discrete dispersion models

We now consider multivariate generalizations of some of the above results, in particular a multivariate Poisson-Tweedie model (cf. Section 5.2). We refer to Johnson *et al.* (1997) for general results on multivariate discrete distributions.

5.1 Multivariate factorial cumulants and other properties

If \mathbf{X} is a k -variate random vector, and \mathbf{s} a k -vector with non-negative elements, we use the notation $\mathbf{s}^{\mathbf{X}} = s_1^{X_1} \cdots s_k^{X_k}$. The multivariate FCGF (Johnson *et al.*, 1997, p. 4) is defined by

$$C(\mathbf{t}; \mathbf{X}) = \log E \left[(1 + \mathbf{t})^{\mathbf{X}} \right] \text{ for } \mathbf{t} \geq -\mathbf{1},$$

where $\mathbf{1}$ is a vector of ones, and the inequality $\mathbf{t} \geq -\mathbf{1}$ is understood elementwise. The effective domain for C is defined by $\text{dom}(C) = \{\mathbf{t} \geq -\mathbf{1} : C(\mathbf{t}) < \infty\}$. When $\mathbf{0} \in \text{int}(\text{dom}(C))$, the mean vector is $E(\mathbf{X}) = \dot{C}(\mathbf{0}; \mathbf{X})$, and the *dispersion matrix* $S(\mathbf{X}) = \ddot{C}(\mathbf{0}; \mathbf{X}) = \text{Cov}(\mathbf{X}) - \text{diag}[E(\mathbf{X})]$ is a $k \times k$ symmetric matrix with entries

$$S_{ij}(\mathbf{X}) = \begin{cases} S(X_i) & \text{for } i = j \\ \text{Cov}(X_i, X_j) & \text{for } i \neq j. \end{cases}$$

We now present a new definition of multivariate over/underdispersion based on the dispersion matrix. We say that the random vector \mathbf{X} is *equidispersed* if $S(\mathbf{X}) = 0$. If \mathbf{X} is not equidispersed, it is called *over/underdispersed* if the dispersion matrix $S(\mathbf{X})$ is positive/negative semidefinite, i.e. $S(\mathbf{X})$ has at least one positive/negative eigenvalue, respectively. We say that the dispersion of \mathbf{X} is *indefinite* if $S(\mathbf{X})$ has both positive and negative eigenvalues.

As an example, consider the bivariate Poisson distribution defined by

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} U_1 + U_2 \\ U_1 + U_3 \end{bmatrix}, \quad (5.1)$$

where $U_i \sim \text{Po}(\mu_i)$, $i = 1, 2, 3$ are independent Poisson random variables. The two marginals X_1 and X_2 are equidispersed, and provided that $\mu_1 > 0$, the marginals are positively correlated, in which case the dispersion is indefinite. In the independence case $\mu_1 = 0$ we find that \mathbf{X} is equidispersed. More generally,

if the marginals of \mathbf{X} are independent and Poisson distributed with mean vector $\boldsymbol{\mu} \geq \mathbf{0}$, we obtain the FCGF

$$C(\mathbf{t}; \mathbf{X}) = \boldsymbol{\mu}^\top \mathbf{t}, \quad (5.2)$$

which is linear, and hence equidispersed. We note in passing, that the multivariate Poisson FCGF (5.2) is of homogeneous type, i.e. of the form $K(\mathbf{t}^\top \boldsymbol{\mu})$, where $K(0) = 0$, see Johnson *et al.* (1997, p. 19). The multinomial distribution $\mathbf{X} \sim \text{Mu}(\boldsymbol{\mu}, n)$ has dispersion matrix $S(\mathbf{X}) = -n\boldsymbol{\mu}\boldsymbol{\mu}^\top$, making this distribution underdispersed.

We now derive the scaling properties of the dispersion matrix $S(\mathbf{X})$ with respect to dilation, generalizing the results of Section 2.2. For a random vector \mathbf{X} , we define the *dilation linear combination* $\mathbf{c} \cdot \mathbf{X}$ with coefficient vector $\mathbf{c} \geq \mathbf{0}$ ($1 \times k$) as follows:

$$C(\mathbf{t}; \mathbf{c} \cdot \mathbf{X}) = C(\mathbf{c}^\top \mathbf{t}; \mathbf{X})$$

provided that the right-hand side is a (univariate) FCGF. The mean and dispersion matrix of a dilation linear combination are given by

$$E(\mathbf{c} \cdot \mathbf{X}) = \mathbf{c}E(\mathbf{X}) \text{ and } S(\mathbf{c} \cdot \mathbf{X}) = \mathbf{c}S(\mathbf{X})\mathbf{c}^\top, \quad (5.3)$$

respectively. It follows that if $\mathbf{c} \cdot \mathbf{X}$ is equidispersed for some $\mathbf{c} \neq \mathbf{0}$, then the dispersion matrix $S(\mathbf{X})$ is singular. The reverse implication holds if the vector $\mathbf{c} \geq \mathbf{0}$ is such that $\mathbf{c}S(\mathbf{X})\mathbf{c}^\top = 0$. Similarly, for an $\ell \times k$ matrix $\mathbf{A} \geq \mathbf{0}$ we define $\mathbf{A} \cdot \mathbf{X}$ by

$$C(\mathbf{t}; \mathbf{A} \cdot \mathbf{X}) = C(\mathbf{A}^\top \mathbf{t}; \mathbf{X}),$$

again provided that the right-hand side is an FCGF. For a multivariate Poisson random vector \mathbf{X} with FCGF (5.2) this yields the following transformation

$$C(\mathbf{t}; \mathbf{A} \cdot \mathbf{X}) = \boldsymbol{\mu}^\top \mathbf{A}^\top \mathbf{t} = (\mathbf{A}\boldsymbol{\mu})^\top \mathbf{t}$$

making $\mathbf{A} \cdot \mathbf{X}$ multivariate Poisson with mean $\mathbf{A}\boldsymbol{\mu}$.

We now turn to a multivariate version of the law of thin numbers. We define the dilation average for the i.i.d. sequence $\mathbf{X}_1, \mathbf{X}_2, \dots$ of $k \times 1$ random vectors by

$$\overline{\mathbf{X}}_n = (n\mathbf{I})^{-1} \cdot \mathbf{S}_n.$$

where $\mathbf{S}_n = \mathbf{X}_1 + \dots + \mathbf{X}_n$ denotes the n th partial sum and \mathbf{I} is the identity matrix. We assume that the \mathbf{X}_i are discrete with mean vector $\boldsymbol{\mu} \geq \mathbf{0}$. Similar to the univariate case in Section 2.4, we obtain the FCGF for $\overline{\mathbf{X}}_n$ as follows:

$$\begin{aligned} C(\mathbf{t}; \overline{\mathbf{X}}_n) &= C((n\mathbf{I})^{-1} \mathbf{t}; \mathbf{X}_1 + \dots + \mathbf{X}_n) \\ &= nC(n^{-1} \mathbf{t}; \mathbf{X}_1) = \boldsymbol{\mu}^\top \mathbf{t} + O(n^{-1}), \end{aligned} \quad (5.4)$$

which converges to the multivariate Poisson FCGF (5.2) as $n \rightarrow \infty$.

To show Hermite convergence, we consider an i.i.d. sequence of discrete random vectors \mathbf{X}_i with $E(\mathbf{X}_1) = \mathbf{m} > \mathbf{0}$ and $S(\mathbf{X}_1) = \boldsymbol{\Sigma} > \mathbf{0}$. Define the translated standardized variable $\mathbf{Z}_n(\boldsymbol{\mu})$ for $\boldsymbol{\mu} \geq \boldsymbol{\Sigma}\mathbf{1}$ by

$$\mathbf{Z}_n(\boldsymbol{\mu}) = (n\mathbf{I})^{-1/2} \cdot \left[\mathbf{S}_n \ominus \left(n\mathbf{m} - n^{1/2}\boldsymbol{\mu} \right) \right],$$

where the Poisson subtraction \ominus is defined by analogy with (2.10). By expanding the FCGF of $\mathbf{Z}_n(\boldsymbol{\mu})$ we obtain

$$C(\mathbf{t}; \mathbf{Z}_n(\boldsymbol{\mu})) = \boldsymbol{\mu}^\top \mathbf{t} + \frac{1}{2} \mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t} + O(n^{-1/2}),$$

which shows that the multivariate Hermite distribution of (cf. Johnson *et al.*, 1997, p. 274) with mean vector $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$ appears in the limit as $n \rightarrow \infty$. Hence $\mathbf{Z}_n(\boldsymbol{\mu})$ converges in distribution to the multivariate Hermite distribution.

Finally, let us consider a multivariate generalization of the zero-inflation index (2.13), namely

$$\text{ZI}(\mathbf{X}) = 1 + \frac{\log P(\mathbf{X} = \mathbf{0})}{E(X_1) + \dots + E(X_k)} = 1 + \frac{C(-\mathbf{1}; \mathbf{X})}{\mathbf{1}^\top \dot{C}(\mathbf{0}; \mathbf{X})}. \quad (5.5)$$

This index measures zero-inflation/deflation relative to independent Poisson random variables (equidispersion) with the same total mean, corresponding to positive/negative values of $\text{ZI}(\mathbf{X})$, respectively. It is useful to extend this to a directional measure of zero-inflation, namely

$$\text{ZI}(\mathbf{c} \cdot \mathbf{X}) = 1 + \frac{C(-\mathbf{1}; \mathbf{c} \cdot \mathbf{X})}{\dot{C}(\mathbf{0}; \mathbf{c} \cdot \mathbf{X})} = 1 + \frac{C(-\mathbf{c}^\top; \mathbf{X})}{\mathbf{c} \dot{C}(\mathbf{0}; \mathbf{X})}$$

which reduces to (5.5) for $\mathbf{c} = \mathbf{1}^\top$. This index measures zero-inflation/deflation for the dilation linear combination $\mathbf{c} \cdot \mathbf{X}$ as a function of \mathbf{c} .

5.2 Multivariate Poisson-Tweedie models

We now introduce a new class of multivariate Poisson-Tweedie mixtures, which is based on the multivariate Tweedie distributions of Jørgensen & Martínez (2013). Consider the k -variate Tweedie distribution $\mathbf{Y} \sim \text{Tw}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with mean vector $\boldsymbol{\mu}$ and covariance matrix

$$\text{Cov}(\mathbf{Y}) = [\boldsymbol{\mu}]^{p/2} \boldsymbol{\Sigma} [\boldsymbol{\mu}]^{p/2} \quad (5.6)$$

where $\boldsymbol{\Sigma}$ denotes a $k \times k$ symmetric positive-definite matrix, and the notation $[\boldsymbol{\mu}]^{p/2}$ denotes a power of the diagonal matrix $[\boldsymbol{\mu}] = \text{diag}(\boldsymbol{\mu})$. By construction, this distribution has univariate Tweedie marginals, see Jørgensen & Martínez (2013).

Let us define the *multivariate Poisson-Tweedie model* $\mathbf{X} \sim \text{PT}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as a Poisson mixture

$$\mathbf{X} | \mathbf{Y} \sim \text{independent Po}(Y_i) \text{ for } i = 1, \dots, k,$$

where X_1, \dots, X_k are assumed conditionally independent given \mathbf{Y} . The multivariate Poisson-Tweedie model has univariate Poisson-Tweedie margins, $X_i \sim \text{PT}_p(\mu_i, \sigma_{ii})$, where σ_{ij} denote the entries of $\boldsymbol{\Sigma}$. The mean vector is $\boldsymbol{\mu}$ and the dispersion matrix is (5.6) (positive-definite) making the distribution overdispersed. The covariance matrix for \mathbf{X} has the form

$$\text{Cov}(\mathbf{X}) = [\boldsymbol{\mu}] + [\boldsymbol{\mu}]^{p/2} \boldsymbol{\Sigma} [\boldsymbol{\mu}]^{p/2},$$

making it straightforward to fit multivariate Poisson-Tweedie regression models using quasi-likelihood. The multivariate Poisson-Tweedie model satisfies the following dilation property:

$$[\mathbf{c}] \cdot \text{PT}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \text{PT}_p([\mathbf{c}] \boldsymbol{\mu}, [\mathbf{c}]^{1-p/2} \boldsymbol{\Sigma} [\mathbf{c}]^{1-p/2}),$$

where \mathbf{c} is a k -vector with positive elements, generalizing the univariate dilation property (3.11). In this way, we obtain multivariate generalizations of all the Poisson-Tweedie models of Table 1 for $p \geq 1$, including multivariate Neyman Type A, Pólya-Aeppli, negative binomial and Poisson-inverse Gaussian distributions.

6 Discussion

In this paper we have developed a new class of discrete factorial dispersion models based on exploring the properties of the factorial cumulant generating function, and we have shown that the dispersion function is a powerful characterization and convergence tool for factorial dispersion models. In particular, the Poisson-Tweedie convergence theorem implies that Poisson-Tweedie models are likely to appear frequently in practice, making these models especially useful for modelling overdispersed count data. These results depend in a crucial way on interpreting the dilation operator as a discrete analogue of scaling.

These results show that factorial dispersion models are in many ways analogous to exponential dispersion models and to the recently proposed classes of extreme and geometric dispersion models (Jørgensen *et al.*, 2010; Jørgensen & Kokonendji, 2011). A common trait for these four types of dispersion models is the role of power asymptotics, which in the extreme dispersion model case implies some of the classical convergence results for extremes towards generalized extreme value distributions (Weibull, Fréchet and Gumbel distributions), see Jørgensen *et al.* (2010) for details. It seems likely that there exist further types of dispersion models with a similar structure, for example in free probability, where Bryc (2009) has introduced so-called free exponential families, and studied an analogue of quadratic variance functions.

Many of our results have multivariate analogues, and in particular we have introduced a class of multivariate Poisson-Tweedie mixtures with Poisson-Tweedie margins. We have introduced a multivariate notion of over- and underdispersion, and a multivariate zero-inflation index. We have also shown that the dilation properties of the dispersion matrix are similar to the scaling properties of the covariance matrix.

There remain a number of further questions to be dealt with for factorial dispersion models. In particular, we need to develop methods for probability calculations and simulations further. We are currently developing methods for quasi-likelihood estimation and inference for multivariate Poisson-Tweedie models, along the same lines as Jørgensen *et al.* (2011). We would also like to obtain a better understanding of underdispersion for factorial dispersion models, perhaps based on the M-transformation, where, however, we are faced with the problem of deciding on the existence of the M-transformation in each case. Finally, it seems possible to obtain new types of point processes based on infinitely divisible factorial dispersion models. In particular, point processes based on Poisson-Tweedie models would seem to have useful dilation properties.

Acknowledgements

We are grateful to Christian Weiß for useful comments on a previous version of the paper.

Appendix A: Exponential dispersion models

In this appendix, we summarize some relevant facts about exponential dispersion models and Tweedie models. An exponential dispersion model $ED(\mu, \gamma)$ with mean $\mu \in \Omega$, dispersion parameter $\gamma > 0$ and unit variance function $V(\mu)$ has PDF of the form

$$f(y; \mu, \gamma) = a(y; \gamma) \exp \left[-\frac{1}{2\gamma} d(y; \mu) \right] \text{ for } y, \mu \in \Omega, \quad (6.1)$$

where the unit deviance function $d(y; \mu)$ is defined by

$$d(y; \mu) = 2 \int_{\mu}^y \frac{y - z}{V(z)} dz \text{ for } y, \mu \in \Omega.$$

The model (6.1) is, for each known value of γ , a natural exponential family with variance function $\gamma V(\mu)$. Hence, the function $a(y; \gamma)$ may be determined by Fourier inversion from the CGF, which may in turn be obtained from V . The model $\text{ED}(\mu, \gamma)$ satisfies the following reproductive property:

$$\bar{Y}_n \sim \text{ED}(\mu, \gamma/n), \quad (6.2)$$

where \bar{Y}_n is the average of Y_1, \dots, Y_n , which are i.i.d. from $\text{ED}(\mu, \gamma)$.

The Tweedie exponential dispersion model $\text{Tw}_p(\mu, \gamma)$ has mean μ and unit variance function

$$V(\mu) = \mu^p \quad \text{for } \mu \in \Omega_p, \quad \text{where } p \notin (0, 1).$$

The domain for μ is either $\Omega_0 = R$ or $\Omega_p = R_+$ for $p \neq 0$. Tweedie models satisfy the scaling property

$$c \text{Tw}_p(\mu, \gamma) = \text{Tw}_p(c\mu, c^{2-p}\gamma) \quad \text{for } c > 0. \quad (6.3)$$

Conventional Tweedie asymptotics (Jørgensen *et al.*, 1994) have the following form. If $\text{ED}(\mu, \gamma)$ with unit variance function $V(\mu)$ satisfies

$$V(\mu) \sim \mu^p \text{ as } \mu \downarrow 0 \text{ or } \mu \rightarrow \infty$$

then

$$c^{-1} \text{ED}(c\mu, c^{2-p}\gamma) \xrightarrow{D} \text{Tw}_p(\mu, \gamma) \quad \text{as } c \downarrow 0 \text{ or } c \rightarrow \infty, \quad (6.4)$$

respectively. The proof is based on convergence of the variance function on the left-hand side of (6.4),

$$c^{-2} c^{2-p} \gamma V(c\mu) \rightarrow \gamma \mu^p,$$

applying Mora's (1990) convergence theorem. The case $c^{2-p} \rightarrow \infty$ requires the model $\text{ED}(\mu, \gamma)$ to be infinite divisible. This result implies a Tweedie approximation, by means of (6.3)

$$\text{ED}(c\mu, c^{2-p}\gamma) \dot{\sim} \text{Tw}_p(c\mu, c^{2-p}\gamma) \quad \text{as } c \downarrow 0 \text{ or } c \rightarrow \infty.$$

In some cases, we have a large-sample interpretation of Tweedie convergence. Let us consider the average \bar{Y}_n with distribution (6.2). Then for $p \neq 2$ we obtain

$$n^{-1/(p-2)} \text{ED}(n^{1/(p-2)} \mu, \gamma/n) \xrightarrow{D} \text{Tw}_p(\mu, \gamma) \quad \text{as } n \rightarrow \infty.$$

We interpret this result as saying that the scaled and exponentially tilted average \bar{Y}_n converges to a Tweedie distribution as $n \rightarrow \infty$.

Appendix B: Proof of Theorem 4.1

Consider a sequence of factorial tilting families $\text{FT}_n(\mu)$ with local dispersion functions v_n having domain Ψ_n and FCGF C_n satisfying the conditions of Theorem 4.1. The idea of the proof is to obtain the FCGF derivative \dot{C} from the limiting dispersion function v , and in turn use the uniform convergence to show convergence of the sequence C_n .

We begin by considering the nonzero case, where $v(\mu) \neq 0$ for $\mu \in \Psi_0$. Let K be a given compact subinterval of Ψ_0 . By assumption $\Psi_0 \subseteq \text{int}(\lim \Psi_n)$, so we may assume that $K \subseteq \Psi_n$ from some n_0 on. We only need to consider $n > n_0$ from now on. Fix a $\mu_0 \in \text{int } K$. Let $\psi_n = \dot{C}_n^{-1}$ denote the inverse FCGF derivative defined by $\dot{\psi}_n(\mu) = 1/v_n(\mu)$ on Ψ_n and $\psi_n(\mu_0) = 0$. Let \dot{C}_n, C_n etc. denote the quantities associated with this parametrization. Similarly, define $\psi : \Psi_0 \rightarrow \mathbb{R}$ by $\dot{\psi}(\mu) = 1/v(\mu)$ on Ψ_0 and $\psi(\mu_0) = 0$. Then for $\mu \in K$

$$\left| \dot{\psi}_n(\mu) - \dot{\psi}(\mu) \right| = \frac{|v_n(\mu) - v(\mu)|}{v_n(\mu)v(\mu)}. \quad (6.5)$$

By the uniform convergence of $v_n(\mu)$ to $v(\mu)$ on K , it follows that $\{v_n(\mu)\}$ is uniformly bounded on K . Since $v(\mu)$ is bounded on K , it follows from (6.5) and from the uniform convergence of v_n that $\dot{\psi}_n(\mu) \rightarrow \dot{\psi}(\mu)$ uniformly on K . This and the fact that $\psi_n(\mu_0) = \psi(\mu_0)$ for all n implies, by a result from Rudin (1976, Theorem 7.17), that $\psi_n(\mu) \rightarrow \psi(\mu)$ uniformly on K . Since K was arbitrary, we have $\psi_n(\mu) \rightarrow \psi(\mu)$ for all $\mu \in \Psi_0$.

Let $I_n = \psi_n(\Psi_n)$ and $I_0 = \psi(\Psi_0) \subseteq \text{int}(\lim I_n)$. Let $J = \psi(K) \subseteq I_0$ and $J_n = \psi_n(K) \subseteq I_n$. Define $\dot{C} : I_0 \rightarrow \Psi_0$ by $\dot{C}(y) = \psi^{-1}(y)$. Since ψ is strictly monotone and differentiable, the same is the case for \dot{C} . Let $\mu \in K$ be given and let $y = \psi(\mu) \in J$ and $y_n = \psi_n(\mu) \in J_n$. Since $v_n(\mu)$ is uniformly bounded on K , there exists an $M > 0$ such that $|v_n(\mu)| \leq M$ for all n and $\mu \in K$. It follows that $|\ddot{C}_n(y)| = |v_n(\dot{C}_n(y))| \leq M$ for all $y \in J$ due to the fact that $J \subseteq J_n$ for n large enough. Since $\mu = \dot{C}(y) = \dot{C}_n(y_n)$ we find, using the mean value theorem, that

$$\begin{aligned} |\dot{C}_n(y) - \dot{C}(y)| &= |\dot{C}_n(y) - \dot{C}_n(y_n)| \\ &\leq M |y - y_n| \\ &= M |\psi(\mu) - \psi_n(\mu)|. \end{aligned}$$

This implies that $\dot{C}_n(y) \rightarrow \dot{C}(y)$ uniformly in $y \in J$. Since $C(0) = C_n(0)$ for all n , it follows by similar arguments as above that $C_n(y) \rightarrow C(y)$ uniformly on J . We conclude from the convergence of the sequence of MGFs $\exp[C_n(e^s - 1)] \rightarrow \exp[C(e^s - 1)]$ for $s \in \log(J + 1)$ that the sequence of distributions $\text{FT}_n(\mu_0)$ converges weakly to a probability measure P with FCGF C . We let $\text{FT}(\mu)$ denote the factorial tilting family generated by P with local dispersion function v on Ψ_0 . We may now complete the proof in the nonzero case by proceeding like in the proof of Proposition 2.1.

In the case where $v(\mu) = 0$ (the zero case), we cannot define the function ψ as above. Instead we take $C(t) = t\mu_0$, such that $\dot{C}(t) = \mu_0$ and $\ddot{C}(t) = 0$ for $t \in \mathbb{R}$. For any $\epsilon > 0$, we may choose an n_0 such that $|v_n(\mu)| \leq \epsilon$ for any $n \geq n_0$ and $\mu \in K$. For such n and μ we hence obtain

$$|\psi_n(\mu)| = \int_{\mu_0}^{\mu} \frac{1}{|v_n(t)|} dt \geq \frac{|\mu - \mu_0|}{\epsilon},$$

which can be made arbitrarily large by choosing ϵ small. We hence conclude that $J_n = \psi_n(K) \rightarrow \mathbb{R}$ as $n \rightarrow \infty$.

Now we let J be a compact interval such that $0 \in \text{int} J$, implying that $J \subseteq J_n$ for n large enough. For such n we hence obtain that $|\ddot{C}_n(t)| = |v_n(\dot{C}_n(t))| \leq \epsilon$ for all $t \in J$, because then $\dot{C}_n(t) \in K$. Since $\mu_0 = \dot{C}(t) = \dot{C}_n(0)$ we find, again by the mean value theorem, that for $t \in J$,

$$|\dot{C}_n(t) - \dot{C}(t)| = |\dot{C}_n(t) - \dot{C}_n(0)| \leq \epsilon |t|.$$

This implies that $\dot{C}_n(t) \rightarrow \dot{C}(t)$ uniformly in $t \in J$. By similar arguments as above, we conclude that $\text{FT}_n(\mu_0)$ converges weakly to a probability measure P with FCGF $C(t) = t\mu_0$, which implies the desired conclusion in the zero case, completing the proof.

References

- BARRETO-SOUZA, W., & BOURGUIGNON, M. 2014. A skew INAR(1) process on \mathbb{Z} . *AStA Advances in Statistical Analysis*, DOI, 10.1007/s10182-014-0236-2.
- BRYC, W. 2009. Free exponential families as kernel families. *Demonstr. Math.*, **XLII**, 657–672.

- DOBBIE, M. J., & WELSH, A. H. 2001. Models for zero-inflated count data using the Neyman type A distribution. *Statistical Modelling*, **1**, 65–80.
- EL-SHAARAWI, A. H., ZHU, R., & JOE, H. 2011. Modelling species abundance using the Poisson-Tweedie family. *Environmetrics*, **22**, 152–164.
- GILES, D. E. 2010. Hermite regression analysis of multi-modal count data. *Economics Bulletin*, **30**, 2936–2945.
- HARREMOËS, P., JOHNSON, O., & KONTOYIANNIS, I. 2010. Thinning, entropy, and the law of thin numbers. *IEEE Transactions on Information Theory*, **56**, 4228–4244.
- JENSEN, S. T., & NIELSEN, B. 1997. On convergence of multivariate Laplace transforms. *Statist. Probab. Lett.*, **33**, 125–128.
- JOHNSON, N. L., KOTZ, S., & BALAKRISHNAN, N. 1997. *Discrete Multivariate Distributions*. New York: Wiley.
- JOHNSON, N. L., KEMP, A. W., & KOTZ, S. 2005. *Univariate Discrete Distributions*. 3rd edn. Hoboken, N.J.: Wiley.
- JØRGENSEN, B. 1997. *The Theory of Dispersion Models*. London: Chapman & Hall.
- JØRGENSEN, B., & KOKONENDJI, C. C. 2011. Dispersion models for geometric sums. *Brazilian J. Probab. Statist.*, **25**, 263–293.
- JØRGENSEN, B., & MARTÍNEZ, J. R. 2013. Multivariate exponential dispersion models. *Pages 73–98 of: KOLLO, T. (ed), Multivariate Statistics: Theory and Applications. Proceedings of the IX Tartu Conference on Multivariate Statistics & XX International Workshop on Matrices and Statistics*. Singapore: World Scientific.
- JØRGENSEN, B., MARTÍNEZ, J. R., & TSAO, M. 1994. Asymptotic behaviour of the variance function. *Scand. J. Statist.*, **21**, 223–243.
- JØRGENSEN, B., GOEGBEUR, Y., & MARTÍNEZ, J. R. 2010. Dispersion models for extremes. *Extremes*, **13**, 399–437.
- JØRGENSEN, B., DEMÉTRIO, C. G. B., KRISTENSEN, E., BANTA, G. T., PETERSEN, H. C., & DELEFOSSE, M. 2011. Bias-corrected Pearson estimating functions for Taylor’s power law applied to benthic macrofauna data. *Statist. Probab. Lett.*, **81**, 749–758.
- KARLIS, D., & XEKALAKI, E. 2005. Mixed Poisson distributions. *Int. Statist Rev.*, **73**, 35–58.
- KEMP, A. W. 1997. Characterizations of a discrete normal distribution. *J. Statist.Plann. Inf.*, **63**, 223–229.
- KEMP, C. D., & KEMP, A. W. 1965. Some properties of the ‘Hermite’ distribution. *Biometrika*, **52**, 381–394.
- KHATRI, C. G. 1959. On certain properties of power-series distributions. *Biometrika*, **46**, 486–490.
- KOKONENDJI, C. C., & PÉREZ-CASANY, M. 2012. A note on weighted count distributions. *Journal of Statistical Theory and Applications*, **11**, 337–352.
- KOKONENDJI, C. C., DOSSOU-GBÉTÉ, S., & DEMÉTRIO, C. G. B. 2004. Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes. *SORT: Statistics and Operations Research Transactions*, **28**, 201–214.

- KOKONENDJI, C. C., MIZÈRE, D., & BALAKRISHNAN, N. 2008. Connections of the Poisson weight function to overdispersion and underdispersion. *J. Statist. Plann. Inf.*, **138**, 1287–1296.
- MASSÉ, J.-C., & THEODORESCU, R. 2005. Neyman Type A distribution revisited. *Statistica Neerlandica*, **59**, 206–213.
- MCKENZIE, E. 1985. Some simple models for discrete variates time series. *Water Resources Bulletin*, **21**, 645–650.
- MORA, M. 1990. La convergence des fonctions variance des familles exponentielles naturelles. *Ann. Fac. Sci. Toulouse (5)*, **11**, 105–120.
- PISTONE, G., & WYNN, H. P. 1999. Finitely generated cumulants. *Statistica Sinica*, **9**, 1029–1052.
- PUIG, P. 2003. Characterizing additively closed discrete models by a property of their maximum likelihood estimators, with an application to generalized Hermite distributions. *J. Amer. Statist. Assoc.*, **98**, 687–692.
- PUIG, P., & BARQUINERO, F. 2011. An application of compound Poisson modelling to biological dosimetry. *Proc. Royal Society A*, **467**, 897–910.
- PUIG, P., & VALERO, J. 2006. Count data distributions: some characterizations with applications. *J. Amer. Statist. Assoc.*, **101**, 332–340.
- PUIG, P., & VALERO, J. 2007. Characterization of count data distributions involving additivity and binomial subsampling. *Bernoulli*, **13**, 544–555.
- RISTIĆ, M. M., BAKOUCH, H. S., & NASTIĆ, A. S. 2009. A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *Journal of Statistical Planning and Inference*, **139**, 2218–2226.
- ROY, D. 2003. The discrete normal distribution. *Communications in Statistics—Theory and Methods*, **32**, 1871–1883.
- RUDIN, W. 1976. *Principles of Mathematical Analysis*. third edn. New York: McGraw-Hill.
- SELLERS, K.F., BORLE, S., & SHMUELI, G. 2012. The COM-Poisson model for count data: a survey of methods and applications. *Applied Stochastic Models in Business and Industry*, **28**, 104–116.
- SHMUELI, G., MINKA, T. P., KADANE, J. P., BORLE, S., & BOATWRIGHT, P. 2005. A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *Applied Statistics*, **54**, 127–142.
- STEUTEL, F. W., & VAN HARN, K. 1979. Discrete analogues of self-decomposability and stability. *Ann. Probab.*, **7**, 893–899.
- TWEEDIE, M. C. K. 1984. An index which distinguishes between some important exponential families. *Pages 579–604 of: GHOSH, J. K., & ROY, J. (eds), Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*. Calcutta: Indian Statistical Institute.
- WEISS, C. H. 2008. Thinning operations for modeling time series of counts—a survey. *ASTA Advances in Statistical Analysis*, **92**, 319–341.
- WILLMOT, G. E. 1987. The Poisson-inverse Gaussian distribution as an alternative to the negative binomial. *Scand. Actuar. J.*, **1987**, 113–127.
- WIMMER, G., & ALTMANN, G. 1999. *Thesaurus of Univariate Discrete Probability Distributions*. Essen: STAMM Verlag.